

NON-PARAMETRIC EMPIRICAL BAYES ESTIMATION

HANS HEDÉN

1975

NON-PARAMETRIC EMPIRICAL BAYES ESTIMATION

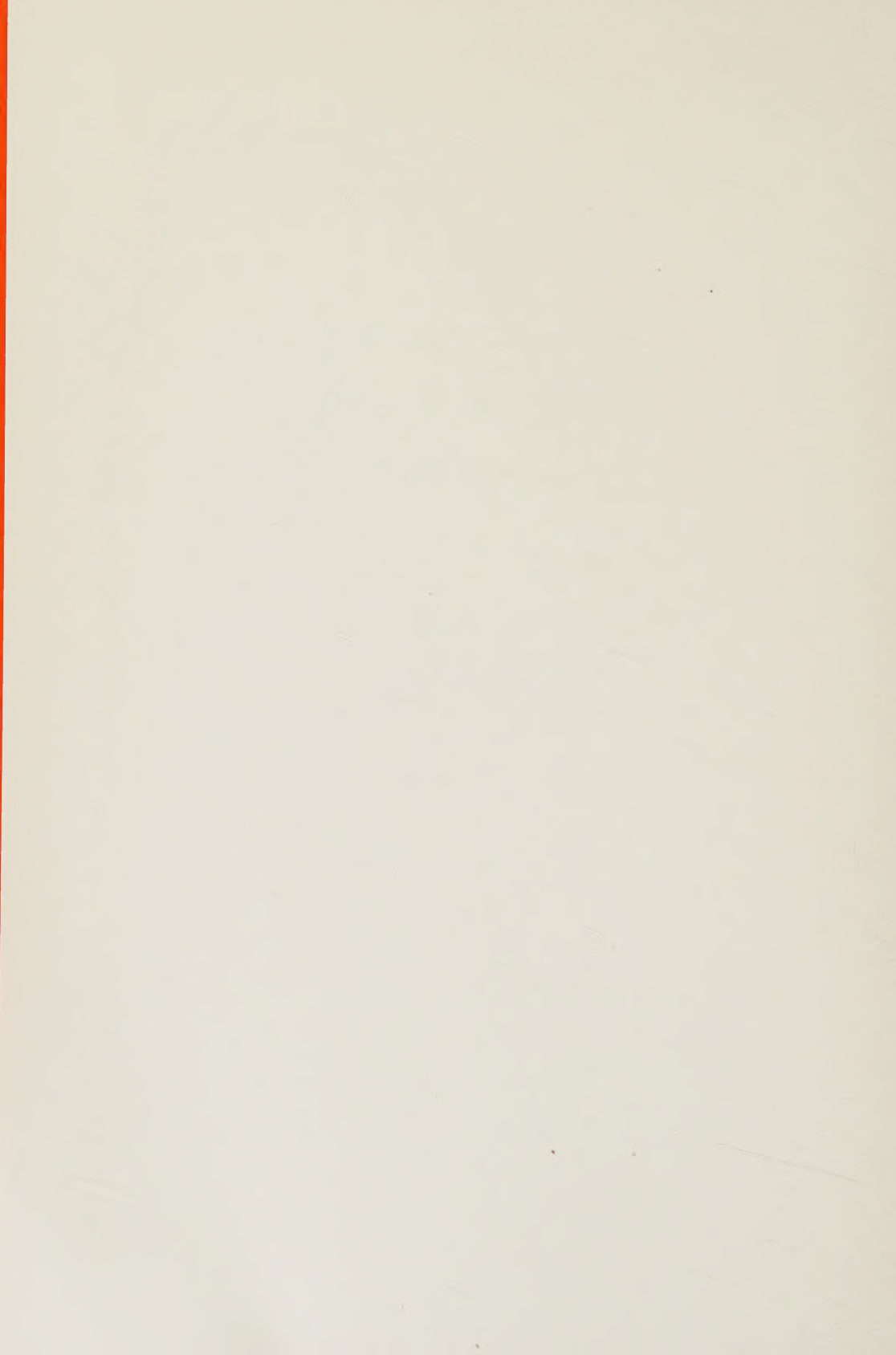
av

Hans Hedén

Fil.kand., Värmlands

AKADEMISK AVHANDLING

som för avläggande av filosofie doktorsexamen vid samhällsvetenskapliga fakulteten vid universitetet i Lund kommer att offentligen försvaras å statistiska institutionen, Getingevägen 8, sal 26, torsdagen den 20 mars 1975, kl 10.15.



**NON-PARAMETRIC
EMPIRICAL BAYES
ESTIMATION**



Digitized by the Internet Archive
in 2026

**NON-PARAMETRIC
EMPIRICAL BAYES
ESTIMATION**

HANS HEDÉN



© Hans Hedén 1975

Printed in Sweden by

Ludvig Larsson Tryckeri AB 1975

ISBN-91-7222-110-0

To BERITH

PREFACE

This book is about Empirical Bayes (EB) estimation, particularly non-parametric estimation. The notion of an EB procedure is due to Robbins (1955). For a time after his original basic work, little was done to develop his ideas and it is only in the last ten years that active research has got under way.

Chapter 1 begins with Bayes and EB methods, and concepts such as loss, risk, extra risk and asymptotic optimality are introduced. In an EB situation, we assume that repeated experiments are carried out, and that for each experiment, an observation or sample is obtained. The samples preceding the current one, will be called past data. It is the parameter underlying the current data which will be estimated.

Chapter 2 reviews a number of EB estimators appearing in the literature. The three non-parametric EB procedures appearing in the literature, which are presented in Johns (1957), Krutchkoff (1967) and Ergle and Krutchkoff (1972) are considered.

Chapter 3 begins by stating that in a non-parametric EB situation, the samples which constitute the current and past data, must contain more than one observation. Let the number of observations in the samples be p . The samples are divided into two parts, containing r and $p-r$ observations respectively. One part is used to construct the Bayes estimator and the other to estimate it. The problem, not considered before, of choosing an optimal r is discussed. New non-parametric EB estimators are presented.

Chapter 4 deals with the non-parametric EB estimators of Chapter 3 in a quality control situation. The calculations of some of the estimators is simplified in this situation. Similarities between different estimators are considered.

Chapter 5 and the Appendix contain the design, discussion and results of a simulation study in the Bernoulli case, carried out by the author. In addition to the non-parametric EB estimators of the thesis, a number of EB estimators proposed by other authors are calculated in the simulation. Thus, it is possible to compare the different estimators. Four solutions to the problem of choosing the r -value are suggested.

As we have already said, Empirical Bayes methods are a rather new topic in Statistics. In Sweden very little research has been done in this field.

During the academic year 1971-72, Ove Frank gave a course on EB methods at the Institute of Statistics at the University of Stockholm. It was during this course that he initiated the research which forms the basis for this monograph.

Ever since I was introduced to the EB field, my teacher Professor Ove Frank, has been invaluable as regards the development of the research project. I thank him for all the stimulating discussions we have had and for the improvements he has suggested in earlier versions of the manuscript.

The table in the Appendix has been arranged by my wife Berith. FL Bengt Swensson and FK Bengt Sundström read the manuscript and gave valuable comments. Craig McKay, M.Sc. improved my English. Earlier versions of the manuscript were typed by Mrs. Caisy Pettersson and Mrs. Ingrid Jonsson. The final text was typed by Mrs. Git Sundt. I thank them all for their able assistance.

Finally, I thank the University of Uppsala for financially supporting the data processing.

Arboga, January 1975

Hans Hedén

CONTENTS

1.	Introduction	11
1.1	The decision problem	11
1.2	Bayes methods	11
1.3	The Bayes estimator and the Bayes risk at quadratic loss	12
1.4	Empirical Bayes methods	13
1.5	Criteria for selecting EB estimators	14
1.6	Extra risk	15
2.	Survey of literature	17
2.1	Some EB procedures	17
2.2	Some non-parametric EB procedures	27
2.3	Design of Monte Carlo simulations of EB procedure	36
3.	Non-parametric EB estimators	39
3.1	Basic concepts	39
3.2	Direct estimation of the Bayes estimator	41
3.2.1	γ procedure	41
3.2.2	ϵ procedure	53
3.3	An example	58
3.4	Estimation of the Bayes estimator with regression procedure	62
3.4.1	ϕ procedure	62
3.4.2	ψ and ω procedures	66
3.4.3	Choice of one estimator among $p-1$ estimators of the same type	68
4.	The quality control problem	74
4.1	Notation	74
4.2	Five non-parametric EB procedures	75
4.2.1	γ procedure	75
4.2.2	ϵ procedure	76
4.2.3	ϕ procedure	86
4.2.4	Identities between γ and ϕ procedures	86
4.2.5	γ and ω procedures	90

5.	Monte Carlo simulation in the Bernoulli case	99
5.1	Design	99
5.2	Procedures introduced in the thesis	101
5.2.1	Choice of r -value and reduction of the number of procedures	101
5.3	Comparison procedures	102
5.4	Result and discussion	111
	Appendix	117
	References	130
	Index	131

1. INTRODUCTION

1.1 The decision problem

The decision problem which will be treated in this work, is point estimation of the parameter θ . An observation, or a vector of observations $\underline{z}$, is generated from the stochastic variable with the distribution function $F(\underline{z}|\theta)$ which depends on θ . The problem is to make a decision $\eta(\underline{z})$ concerning the value of θ . A decision function $\eta(\underline{z})$, with acceptable properties, is then selected from the set of all conceivable decision functions.

In the situation considered, repeated experiments are performed. To every experiment, there corresponds a parameter θ , which will be estimated. For every experiment, a sample $\underline{z}_i$ is obtained from the variable with the distribution function $F(\underline{z}|\theta_i)$.

1.2 Bayes methods

In the Bayesian approach to the problem of estimating θ , the parameter θ is regarded as a realisation of a stochastic variable with the distribution function $G(\theta)$. The distribution $G(\theta)$ is called the prior distribution of the parameter.

Bayes solution depends on $G(\theta)$ and Bayes decision function is denoted by $\delta(\underline{z})$. To be able to derive Bayes decision function, $\delta(\underline{z})$, a loss function must be defined. Let the loss, $L(\eta(\underline{z}), \theta) \geq 0$, be incurred when the parameter is θ and the decision $\eta(\underline{z})$ is made. In this work, where $\eta(\underline{z})$ is a point estimator of the parameter θ , it is usual and practical to let the loss function be a quadratic loss function

$$L(\eta, \theta) = (\eta - \theta)^2. \quad (1)$$

To every decision function, $\eta(\underline{z})$, there corresponds a risk function, $R_\eta(\theta)$. If the decision is $\eta(\underline{z})$ and the value of the parameter is θ the risk function is defined as

$$R_\eta(\theta) = E_{\underline{z}|\theta} L(\eta(\underline{z}), \theta). \quad (2)$$

One way to select a decision function is to select the decision function, $\eta(\underline{z})$, whose risk, $R_{\eta}(\theta)$, is as little as possible. It is well known, however, that it is impossible to generally find a decision function, $\eta'(\underline{z})$, such that $R_{\eta'}(\theta) \leq R_{\eta}(\theta)$. Thus, it is impossible to find a decision function $\eta'(\underline{z})$ such that the inequality holds for all θ and η . On the other hand it is possible to find a decision function, $\eta(\underline{z})$, which minimizes the expected risk, $W(\eta)$, derivable with the aid of $G(\theta)$. The expected risk is given by

$$W(\eta) = \int_{\underline{z}, \theta} L(\eta(\underline{z}), \theta) \cdot \quad (3)$$

Bayes method of finding a decision function is to minimize the expected risk, $W(\eta)$. The special decision function, which minimizes the expected risk, is called Bayes decision function. It is denoted by $\delta(\underline{z})$, which has already been mentioned. The expected risk of the Bayes decision function, $\delta(\underline{z})$, is called Bayes risk and is denoted by $W(\delta)$.

1.3 The Bayes estimator and the Bayes risk at quadratic loss

The Bayes estimator is derived by minimizing the expected risk, $W(\eta)$.

Now $W(\eta)$ is

$$W(\eta) = \int_{\underline{z}, \theta} (\eta(\underline{z}) - \theta)^2 \cdot \int f(\underline{z}) \{ \int (\theta - \eta(\underline{z}))^2 g(\theta | \underline{z}) d\theta \} d\underline{z} . \quad (1)$$

The expected risk is minimized when the expression within the braces is minimized with regard to $\eta(\underline{z})$. This minimum is achieved when the decision function or estimator, $\eta(\underline{z})$, equals the expected value of the posterior distribution. In the case of the quadratic loss function the Bayes estimator becomes

$$\delta(\underline{z}) = E(\theta | \underline{z}) . \quad (2)$$

The density function, $g(\theta | \underline{z})$, which is used above, is the density function for θ given $\underline{z}$, deduced from Bayes theorem. It is called the density function of the posterior distribution and is defined by

$$g(\underline{\theta}|\underline{z}) = \frac{f(\underline{z}|\underline{\theta}) g(\underline{\theta})}{\int f(\underline{z}|\underline{\theta}) g(\underline{\theta}) d\underline{\theta}} \quad (3)$$

where $g(\underline{\theta})$ is the density function of the prior distribution.

The Bayes risk is the expected risk which occurs when the Bayes estimator is used. The expected risk or average squared error can be written as

$$\begin{aligned} W(\underline{\eta}) &= E_{\underline{z}, \underline{\theta}} (\underline{\eta}(\underline{z}) - \underline{\theta})^2 = E_{\underline{z}} E_{\underline{\theta}|\underline{z}} (\underline{\eta}(\underline{z}) - \underline{\theta})^2 = \\ &= E_{\underline{z}} (\underline{\eta}(\underline{z}) - E(\underline{\theta}|\underline{z}))^2 + E_{\underline{z}} \text{Var}(\underline{\theta}|\underline{z}) . \end{aligned} \quad (4)$$

Put the estimator $\underline{\eta}(\underline{z})$ equal to the Bayes estimator

$$\underline{\delta}(\underline{z}) = E(\underline{\theta}|\underline{z}) \quad (5)$$

and the Bayes risk, $W(\underline{\delta})$, is

$$W(\underline{\delta}) = E_{\underline{z}} \{ \text{Var}(\underline{\theta}|\underline{z}) \} \quad (6)$$

where $\text{Var}(\underline{\theta}|\underline{z})$ is the variance of the posterior distribution.

A general source of reference on the above matters is Lehmann (1959).

1.4 Empirical Bayes methods

For the Bayes estimator, $\underline{\delta}(\underline{z}) = E(\underline{\theta}|\underline{z})$, to be derivable, the prior distribution, $G(\underline{\theta})$, must be known. If $G(\underline{\theta})$ is unknown or contains unknown parameters, $G(\underline{\theta})$ or the parameters have to be estimated before Bayes method can be used. A run of independent experiments is the base of the estimation. In experiment number i , the parameter has the value θ_i and in the experiment the observation $\underline{z}_i$ is generated. When $n+1$ experiments have been carried out $n+1$ parameter values, $\theta_0, \theta_1, \dots, \theta_n$, are obtained. These parameter values are assumed to be unknown. To these parameter values, correspond the observations, $\underline{z}_0, \underline{z}_1, \dots, \underline{z}_n$. The variables $(\underline{z}_i, \theta_i)$ for $i = 0, 1, \dots, n$ will be stochastically independent and will have the same distribution. The observations, $\underline{z}_1, \dots, \underline{z}_n$, will be called past data

and $\underline{z}_0$ will be called current data. The experiments will be numbered inversely so that the i th experiment is the i th experiment before the current one. Both the prior distribution, $G(\theta)$, and the conditional distribution, $f(\underline{z}|\theta)$, have to be known before the Bayes estimator, $\delta(\underline{z}_0)$ can be constructed. The quantity $\delta(\underline{z}_0)$ is the Bayes estimator of the parameter value of θ_0 of the current experiment. The prior distribution, $G(\theta)$, is not known but by employing past data it is possible to construct an estimate, $\pi(\underline{z}_0, \underline{z}_1, \dots, \underline{z}_n)$, of $\delta(\underline{z}_0)$.

The estimate of the Bayes estimator, which employs past data, can be constructed in two ways. One way is to estimate the prior distribution, $G(\theta)$, and with the aid of this compute the estimate π . Here, however, the problem arises that the distribution $G(\theta)$ which is being estimated has to be identifiable i.e. that $G(\theta)$ can be determined uniquely from the function

$$f(\underline{z}) = \int f(\underline{z}|\theta) dG(\theta) .$$

Identifiability is treated by Maritz (1970). The other way is to estimate the Bayes estimator directly with some method. In this way the problem of identifiability is avoided. The second method will be used in this work.

A procedure which is a function of past as well as current data is called an Empirical Bayes (EB) procedure. The notion of an Empirical Bayes procedure was invented by Robbins (1955).

1.5 Criteria for selecting EB estimators

The Bayes estimator, δ , is the decision function which minimizes the expected risk, W . The expected risk of an arbitrary decision function can be seen as a measure of the inaccuracy of the decision function. In particular the EB estimator, $\pi(\underline{z}_0, \underline{z}_1, \dots, \underline{z}_n)$, has an expected risk, $W(\pi)$, which is a function depending on past data. Thus, $W(\pi)$ will vary randomly from one set of past data to another. The overall expected loss with respect to all variables is

$$E_{\underline{z}_1, \dots, \underline{z}_n} W(\pi) . \quad (1)$$

This is the expected value in the distribution of $W(\pi)$ when there are n past experiments. If the distribution of $W(\pi)$ is known the probability $P(W(\pi) > W(T))$ for

example can be determined, where $W(T)$ is the expected risk of a traditional estimator. T can be the maximum likelihood estimator for example. Another measure of interest when evaluating π , can be obtained by constructing the ratio

$$R = \frac{E \quad W(\pi)}{W(T)} \quad (2)$$

The estimate of the ratio, R , will be used in Chapter 5 to evaluate the different procedures. It is natural to compare EB estimators with traditional estimators because estimators of both kinds can be determined in an EB situation.

Neither the distribution of $W(\pi)$ nor its expected value can be determined theoretically, except possibly for a very simple example. In order to evaluate an EB procedure π other ways must therefore be used. One possibility is to use Monte Carlo simulations and from these estimate the distribution of $W(\pi)$ or its expected value. Another possibility is to analyze if the expected value of $W(\pi)$ converges to $W(\delta)$ when $n \rightarrow \infty$. If

$$\lim_{n \rightarrow \infty} E \quad W(\pi) \rightarrow W(\delta)$$

π is said to be asymptotically optimal cf. Robbins (1964).

1.6 Extra risk

Different decision functions can be compared with the aid of the extra risks of the decision functions. The extra risk, $r(\eta)$, of the decision function, η , is the risk which has to be added to the minimal risk, the Bayes risk, so that the expected risk, $W(\eta)$, will be obtained.

If

$$E(\theta)^2 < \infty \quad (1)$$

it follows that the Bayes risk $W(\delta)$ is finite

$$W(\delta) = E \quad \text{Var}\{\theta | \underline{z}\} < \infty . \quad (2)$$

The extra risk can be expressed as

$$r(\eta) = W(\eta) - W(\delta) = E_{\underline{z}} \{ \eta(\underline{z}) - E(\theta | \underline{z}) \}^2. \quad (3)$$

The extra risk is useful when the asymptotical properties of an EB estimator have to be examined. The EB estimator π is asymptotically optimal if

$$\lim_{n \rightarrow \infty} E_{\underline{z}_1, \dots, \underline{z}_n} r(\pi) = \quad (4)$$

$$= \lim_{n \rightarrow \infty} E_{\underline{z}_0, \underline{z}_1, \dots, \underline{z}_n} \{ \pi(\underline{z}_0, \underline{z}_1, \dots, \underline{z}_n) - E(\theta | \underline{z}_0) \}^2 = 0.$$

2. SURVEY OF LITERATURE

The criterion for the selection of literature is that the EB procedure which is treated has one out of two qualities. The qualities are, the procedure is non-parametric and the procedure is applicable to the binomial distribution.

2.1 Some EB procedures

Robbins (1955), who invented the notion EB, has suggested an EB method for estimating the current value of θ . By using the method we avoid estimating the prior distribution explicitly. The method is applicable to some special types of distributions. The author derives his EB estimator for among other the Poisson and the binomial distribution.

Robbins starts by introducing a discrete stochastic variable z whose distribution function, in a known way, depends on the unknown parameter θ . The probability function is denoted by

$$f(z|\theta) \quad (1)$$

where θ also is a stochastic variable with the distribution function

$$G(\theta) \quad (2)$$

The unconditional distribution of z is denoted by

$$f_G(z) = \int f(z|\theta) dG(\theta) \quad (3)$$

The Bayes estimator at quadratic loss is (cf. Chapter 1.3)

$$\delta(z) = E(\theta|z) = \frac{\int \theta f(z|\theta) dG(\theta)}{\int f(z|\theta) dG(\theta)} \quad (4)$$

Current observation and past observations on z are denoted by $z_0, z_1, \dots, z_n$. Where there is no danger of confusion the current observation will mostly be denoted by just z , in this Chapter.

Let us assume that $f(z | \theta)$ is a Poisson distribution

$$f(z | \theta) = e^{-\theta} \frac{\theta^z}{z!} \quad z = 0, 1, \dots; \theta > 0. \quad (5)$$

The prior distribution G belongs to the class of all distribution functions defined on the positive real axis.

We can express the unconditional distribution of z as

$$f_G(z) = \int_0^\infty f(z | \theta) dG(\theta) = \int_0^\infty e^{-\theta} \theta^z dG(\theta) / z! \quad (6)$$

The Bayes estimator is given by

$$\delta(z) = \frac{\int_0^\infty e^{-\theta} \theta^{z+1} dG(\theta)}{\int_0^\infty e^{-\theta} \theta^z dG(\theta)} \quad (7)$$

The Bayes estimator can be expressed by means of the unconditional distribution of z in the following way

$$\delta(z) = (z + 1) \frac{f_G(z+1)}{f_G(z)} \quad (8)$$

The set of past data $z_1, \dots, z_n$ provides direct estimates of $f_G(z)$ and $f_G(z+1)$. Let $f_n(z)$ denote the number of past observations which are equal to z . The relative frequency $(f_n(z) + 1)/(n + 1)$ is an estimate of $f_G(z)$ if the current observation is included. The corresponding estimate of $f_G(z+1)$ is $f_n(z+1)/(n + 1)$, where $f_n(z+1)$ stands for the number of past observations which are equal to $z + 1$.

Robbins constructs the function

$$\pi(z) = (z + 1) \frac{f_n(z+1)}{1 + f_n(z)} \quad (9)$$

and states that for arbitrary G we have for any fixed z

$$\pi(z) \rightarrow \delta(z) \text{ with probability } 1 \text{ when } n \rightarrow \infty. \quad (10)$$

This, the author writes, suggests using the quantity (9) as an estimate of the unknown current value of θ in the hope that

$$E \{ \pi(z) - \theta \}^2 \rightarrow E \{ \delta(z) - \theta \}^2 \text{ when } n \rightarrow \infty. \quad (11)$$

Robbins never says if (11) is true and thus if the estimator (9) is asymptotically optimal or not.

In the article, the correspondence with (9) is deduced given the binomial distribution. Assume that p represents the number of trials, z the number of successes and θ the unknown probability of success in each trial. Then the conditional distribution is given by

$$f_p(z|\theta) = \binom{p}{z} \theta^z (1-\theta)^{p-z} \quad z = 0, 1, \dots, p; \quad 0 \leq \theta \leq 1. \quad (12)$$

The prior distribution, G , belongs to the class of all distribution functions on the interval $(0, 1)$. The unconditional distribution of z is given by

$$f_{G,p}(z) = \binom{p}{z} \int_0^1 \theta^z (1-\theta)^{p-z} dG(\theta). \quad (13)$$

The Bayes estimator can be expressed as

$$\delta_p(z) = \frac{\int_0^1 \theta^{z-1} (1-\theta)^{p-z} dG(\theta)}{\int_0^1 \theta^z (1-\theta)^{p-z} dG(\theta)}. \quad (14)$$

By using the distribution in (13) and noticing that

$$\begin{aligned}
 f_{G,p+1}^{(z+1)} &= \binom{p+1}{z+1} \int_0^1 \theta^{z+1} (1-\theta)^{p-z} dG(\theta) = \\
 &= \frac{p+1}{z+1} \binom{p}{z} \int_0^1 \theta^{z+1} (1-\theta)^{p-z} dG(\theta)
 \end{aligned} \tag{15}$$

the Bayes estimator can be written as

$$\delta_p(z) = \frac{z+1}{p+1} \frac{f_{G,p+1}^{(z+1)}}{f_{G,p}^{(z)}} \quad z = 1, \dots, p. \tag{16}$$

The distribution $f_{G,p+1}^{(z+1)}$ cannot be estimated. The author constructs instead the estimator

$$\delta_{p-1}(z) = \frac{z+1}{p} \frac{f_{G,p}^{(z+1)}}{f_{G,p-1}^{(z)}} \tag{17}$$

where $f_{G,p}^{(z+1)}$ is estimated by using the number of observations which are equal to $z+1$. Let $f_{n+1,p}^{(z+1)}$ denote this number. In the estimation of $f_{G,p-1}^{(z)}$, the observations are built up as the number of successes in for example the first $p-1$ of the p trials which are performed for every θ -value. The EB estimator of (17) is then

$$\pi_p(z) = \frac{z+1}{p} \frac{f_{n+1,p}^{(z+1)}}{f_{n+1,p-1}^{(z)}}. \tag{18}$$

Robbins points out that it is then the case that

$$\pi_p(z) \rightarrow \delta_{p-1}(z) \text{ with probability 1 when } n \rightarrow \infty. \tag{19}$$

Rutherford and Krutchkoff (1969) show how asymptotically optimal EB estimators are obtained for four classes of distributions. It is shown that the estimator (9) is asymptotically optimal. The authors point out, no asymptotical-

ly optimal estimator has yet been found when the binomial distribution is present.

Maritz (1966) gives an example to show that the Robbins estimator in the Poisson case (9) is worse than conventional estimators, unless n is reasonably large. This, Maritz writes, is not surprising because the estimator (9) depends only on the ratio of two successive frequencies in an observed discrete frequency distribution, the overall form of the distribution being disregarded. Individual observed relative frequencies can fluctuate widely when a small number of past observations is present. The author writes that a natural question is whether better use of the observed distribution would lead to better EB estimators. Maritz suggests some sort of 'smoothing' of $\pi(z)$. The main principle which is considered in the article is estimation of the prior distribution with a step function by using past data. The prior distribution, G , is then changed for its estimate in the Bayes estimator and an EB estimator is given. This estimator is smooth, Maritz (1967) writes, in the sense that it depends on the actual numerical values of all past observations, contrary to the Robbins EB estimators in which not all numerical values in past data are used explicitly.

Lemon and Krutchkoff (1969) have proposed a smooth EB estimator which amounts to estimating the prior distribution $G(\theta)$ by means of its empirical distribution function. The Bayes estimator is then

$$\delta(z) = \frac{\sum_{i=0}^n \theta_i f(z|\theta_i)}{\sum_{i=0}^n f(z|\theta_i)} . \quad (20)$$

If θ_i are estimated with conventional methods in the Poisson case which is the case treated in the article, the estimates $\hat{\theta}_i$ equal to z_i will be used. The EB estimator is then

$$\pi_1(z) = \frac{\sum_{i=0}^n e^{-z_i} z_i^{z+1}}{\sum_{i=0}^n e^{-z_i} z_i^z} . \quad (21)$$

The authors also suggest that the procedure is iterated once. If the estimates $\hat{\theta}_i$ are equal to $\pi_1(z_i)$ we get

$$\pi_2(z) = \frac{\sum_{i=0}^n \exp\{-\pi_1(z_i)\} \{\pi_1(z_i)\}^{z+1}}{\sum_{i=0}^n \exp\{-\pi_1(z_i)\} \{\pi_1(z_i)\}^z} . \quad (22)$$

Lemon and Krutchkoff have compared the average squared error for the EB estimators (21) and (22) with the average squared error for the classical estimator by using Monte Carlo simulations. They have then found that (22) has smaller average squared error than (21) but both of them have a smaller average squared error than the classical estimator. Asymptotical properties are not treated in the article.

Griffin and Krutchkoff (1971a) use, without mentioning it, Robbins' technique to construct the Bayes estimator in the case of binomial distribution. The binomial distribution $f_p(z|\theta)$ has the parameters (p, θ) and the loss function which the authors make use of is defined as

$$L(\eta, \theta) = \frac{(\eta - \theta)^2}{\theta(1 - \theta)} .$$

The Bayes estimator is, if the notations in (13) are used, equal to

$$\delta_p(z) = \frac{z}{p-1} \frac{f_{G,p-1}(z)}{f_{G,p-2}(z-1)} \quad z \neq 0, p . \quad (23)$$

In order to construct an EB estimator, $f_{G,p-1}(z)$ and $f_{G,p-2}(z-1)$ can be estimated with the method which has been proposed by Robbins. Although convergent estimators for $f_{G,p-1}(z)$ and $f_{G,p-2}(z-1)$ can be developed, we forego the asymptotical properties, the authors write. They instead propose the type of smooth mass function estimators similar to the ones described in Lemon and Krutchkoff (1969). The following estimators of the marginal probability functions are employed

$$\hat{f}_{G,p-1}(z) = \frac{1}{n+1} \sum_{i=0}^n \binom{p-1}{z} \hat{\theta}_i^z (1 - \hat{\theta}_i)^{p-z-1} \quad z \neq 0, p \quad (24)$$

$$\hat{f}_{G,p-2}(z) = \frac{1}{n+1} \sum_{i=0}^n \binom{p-2}{z-1} \hat{\theta}_i^{z-1} (1 - \hat{\theta}_i)^{p-z-1} \quad z \neq 0, p \quad (25)$$

The quantity $\hat{\theta}_i$ above is an estimate of the parameter underlying the i th experiment. The authors use the maximum likelihood (ML) estimates $\hat{\theta}_i = z_i/p$ for the $\hat{\theta}_i$'s in (24) and (25) and express an EB estimator of (23) in the following way

$$\pi_p(z) = \begin{cases} 0 & \text{if } z = 0 \\ \frac{z \sum_{i=0}^n \binom{p-1}{z} \left(\frac{z_i}{p}\right)^z \left(1 - \frac{z_i}{p}\right)^{p-z-1}}{(p-1) \sum_{i=0}^n \binom{p-2}{z-1} \left(\frac{z_i}{p}\right)^{z-1} \left(1 - \frac{z_i}{p}\right)^{p-z-1}} & \text{if } z \neq 0, p \\ 1 & \text{if } z = p \end{cases} \quad (26)$$

The expression can be reduced and we get

$$\pi_p(z) = \begin{cases} 0 & \text{if } z = 0 \\ \frac{\sum_{i=0}^n \left(\frac{z_i}{p}\right)^z \left(1 - \frac{z_i}{p}\right)^{p-z-1}}{\sum_{i=0}^n \left(\frac{z_i}{p}\right)^{z-1} \left(1 - \frac{z_i}{p}\right)^{p-z-1}} & \text{if } z \neq 0, p \\ 1 & \text{if } z = p \end{cases} \quad (27)$$

In the article, Monte Carlo simulations are reported, in which the authors have compared the estimated expected risks for (26) and the ML estimator. A table in the article lists the 32 different priors which have been employed in the simulation project. An additional table gives the estimates of the ratio R defined in (1.5 (2)) for $n = 1, 5, 10, 25, 50, 100$ for each prior and for each of three different sample sizes, $p = 5, 10, 25$. The estimator T in (1.5 (2)) is here the ML estimator. They establish that the Monte Carlo investigation demonstrates that even for n as small as 5, the estimator (26) can be used in place of the ML estimator.

Griffin and Krutchkoff (1971b) state that current and past data $(z_0, z_1, \dots, z_n)$ are employed to construct EB estimators of the Bayes estimator $E(\theta|z)$. For certain conditional distributions, $E(\theta|z)$ cannot be consistently estimated unless other assumptions are made. Consequently, the authors write, we restrict our attention to linear estimators $\eta(x) = \alpha + \beta x$ where $x = u(z)$ and $E\{u(z)|\theta\} = \theta$. Within this class of linear estimators, the minimal mean squared error estimator depends only upon the first two prior moments, both of which can often be estimated with $(z_0, z_1, \dots, z_n)$.

The Bayes estimator of linear estimators $\eta(x) = \alpha + \beta x$ is derived, at quadratic loss, provided that the first two moments of both F and G exist. Besides it must be true that $E(x|\theta) = \theta$. To give the smallest expected risk β and α are expressed as

$$\beta = \frac{\text{Var}(\theta)}{\text{Var}(x)} \quad (28)$$

$$\alpha = (1 - \beta) E(\theta) . \quad (29)$$

The Bayes estimator can thus be written as

$$\delta(x) = \frac{\text{Var}(\theta)}{\text{Var}(x)} \{x - E(\theta)\} + E(\theta) . \quad (30)$$

In the case of the binomial distribution, which is treated in the article, we can choose x equal to z/p , if the notations in (12) are used. Consequently

$$\beta = \frac{p^2 \text{Var}(\theta)}{\text{Var}(z)} = \frac{p(E(\theta^2) - E^2(\theta))}{(p-1)E(\theta^2) + E(\theta) - pE^2(\theta)} \quad (31)$$

$$\alpha = (1 - \beta) E(\theta) \quad (32)$$

since

$$\text{Var}(z) = E \underset{\theta}{\text{Var}(z|\theta)} + \text{Var} \underset{\theta}{E(z|\theta)} .$$

By observing that the following identities are true

$$E(\theta) = E_{\theta} E_{z|\theta} (z/p|\theta) \quad (33)$$

$$E(\theta^2) = E_{\theta} E_{z|\theta} \{(z^2 - z)/(p^2 - p)\} \quad (34)$$

the first two prior moments $E(\theta)$ and $E(\theta^2)$ can be estimated. Their moment estimators are

$$m_1 = \frac{1}{n+1} \sum_{i=0}^n \frac{z_i}{p} \quad (35)$$

$$m_2 = \frac{1}{n+1} \sum_{i=0}^n \frac{z_i(z_i - 1)}{p(p-1)} \quad (36)$$

Accordingly, the authors write, we estimate β and α with

$$\beta_{n+1} = \begin{cases} 1 & \text{if } (p-1)m_2 + m_1 - pm_1^2 \leq 0 \text{ or if } m_2 - m_1^2 \leq 0 \\ \frac{p(m_2 - m_1^2)}{(p-1)m_2 + m_1 - pm_1^2} & \text{otherwise} \end{cases} \quad (37)$$

$$\alpha_{n+1} = (1 - \beta_{n+1})m_1 \quad (38)$$

and a linear EB estimator

$$\pi(z/p) = \alpha_{n+1} + \beta_{n+1}(z/p) \quad (39)$$

is obtained.

In the article it is mentioned that the expected risks of the EB estimator (39) and the ML estimator have been compared with the aid of Monte Carlo simulations.

The authors write that, for every combination of p (5, 10 or 25) and prior (96 combinations in all) the estimator $\pi(z/p)$ had a smaller estimated expected risk than the ML estimator if $n \geq 5$.

Bennett and Martz (1972) suggest, that in the Bayes estimator

$$\delta(\underline{z}) = \frac{\int \theta f(\underline{z} | \theta) g(\theta) d\theta}{\int f(\underline{z} | \theta) g(\theta) d\theta} \quad (40)$$

the prior density function $g(\theta)$ is estimated with a density estimator given by Parzen (1962). Parzen's density estimator has earlier been used in an EB situation by Clemmer and Krutchkoff (1968). In past and current samples θ_i are estimated. These estimates are then used in Parzen's density estimator. Bennett and Martz point out that the EB estimator obtained is not asymptotically optimal.

The authors suggest an additional EB estimator, where the density estimator is calculated on transformed data.

In the article it is mentioned that a Monte Carlo simulation has been worked out, where the average squared error of the two suggested EB estimators are compared with the average squared error for the ML estimator. The investigation has been performed for the case that $f(z | \theta)$ is a Poisson distribution. The authors state that the two EB estimators have a smaller average squared error than the ML estimator in all the cases observed. The authors have also compared the two estimators with EB estimators suggested by Maritz (1966) and Lemon and Krutchkoff (1969). The EB estimator, where the estimator of the density function is based on transformed data, is shown to be better than the original EB estimator of the article in the cases considered. In addition it is better than the estimators which have been suggested by Maritz (1966) and Lemon and Krutchkoff (1969), who use a discrete smooth method in order to estimate the prior distribution.

A method of constructing EB estimators is to postulate a certain prior distribution, usually the naturally conjugate distribution cf. Raiffa and Schlaifer (1961). The prior parameters, which will be included in the Bayes estimator, can then be estimated with the past data. Maritz (1970) has reported several different cases.

If prior distribution, $G(\theta)$, is Beta (e, f) and the conditional distribution is binomial (p, θ) the Bayes estimator is

$$\delta(z) = \frac{e + z}{p + e + f} \quad (41)$$

If e and f are estimated with the method of moments we get according to Maritz (1970, p. 53)

$$\hat{e} = \max\{m_1(m_2 - pm_1)/(m_1^2(p-1) + p(m_1 - m_2)), o(n^{-1})\} \quad (42)$$

$$\hat{f} = \max\{(p - m_1)(m_2 - pm_1)/(m_1^2(p-1) + p(m_1 - m_2)), o(n^{-1})\} \quad (43)$$

where m_1 is the first sample moment

$$m_1 = \sum_{i=0}^n z_i / (n+1) \quad (44)$$

and m_2 is the second sample moment

$$m_2 = \sum_{i=0}^n z_i^2 / (n+1) \quad (45)$$

2.2 Non-parametric EB procedures

Johns (1957) has developed a non-parametric EB procedure. He defines past data as samples of p observations each, where $p > 1$. Thus, past data are composed of the p -vectors $\underline{z}_1, \dots, \underline{z}_n$. The observations in each sample are independently generated from the distributions $F(z_{ij} | \theta_i)$ for $j = 1, \dots, p$; $i = 1, \dots, n$. This distribution is not necessarily known.

Let the function u be such that

$$E\{u(z) | \theta\} = \theta \quad (1)$$

and

$$E u^2(z) < \infty . \quad (2)$$

Current data consist of $p-1$ observations and are denoted by $\underline{z}^{(p)}$. The notation $\underline{z}^{(p)}$ denotes the vector which remains when the p th element of the p -vector $\underline{z}$ has been omitted. Johns constructs the estimator

$$\lambda(\underline{z}^{(p)}) = \begin{cases} \frac{\sum_{i=1}^n u(z_{ip}) J(\underline{z}_i^{(p)} = \underline{z}^{(p)})}{\sum_{i=1}^n J(\underline{z}_i^{(p)} = \underline{z}^{(p)})} & \text{if } \sum_{i=1}^n J(\underline{z}_i^{(p)} = \underline{z}^{(p)}) > 0 \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

where

$$J(\underline{z}_i^{(p)} = \underline{z}^{(p)}) = \begin{cases} 1 & \text{if any of the permutations of } (z_{i1}, \dots, z_{i,p-1})' \\ & \text{equals } (z_1, \dots, z_{p-1})' \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

Johns has proved that the estimator (3) is asymptotically optimal both for discrete and continuous distributions $F(z|\theta)$.

If the current sample consist of p observations $\underline{z}$, Johns suggests that all observations in $\underline{z}$ should be used. This will be done by constructing an average of p estimators of the type (3) as

$$\lambda_p = \frac{1}{p} \sum_{j=1}^p \lambda(\underline{z}^{(j)}) \quad (5)$$

where $\underline{z}^{(j)}$ denotes $\underline{z}$ with the j th element omitted.

It holds true that λ_p in (5) is asymptotically optimal if the Bayes estimator is defined as

$$\sum_{j=1}^p \delta(\underline{z}^{(j)})/p . \quad (7)$$

It is not usually true that

$$\delta(\underline{z}) = \sum_{j=1}^p \delta(\underline{z}^{(j)})/p . \quad (8)$$

Thus, λ_p is not asymptotically optimal in the usual sense. Johns has proved that between the expected risks of the estimators (3) and (5) the following inequality can be obtained

$$E_{\underline{z}_1, \dots, \underline{z}_n} W(\lambda_p) \leq E_{\underline{z}_1, \dots, \underline{z}_n} W(\lambda(\underline{z}^{(p)})) . \quad (9)$$

Concerning the risks of the Bayes estimators the following inequality can be obtained

$$W(\sum_{j=1}^p \delta(\underline{z}^{(j)})/p) \leq W(\delta(\underline{z}^{(p)})) . \quad (10)$$

The estimator (5) should then be preferred to the estimator (3). In addition the limit of the estimator (5), as n grows, is a Bayes estimator, whose expected risk is smaller than or equal to the expected risk of the Bayes estimator, which the estimator (3) converges to.

In the article, upper and lower bounds of the expected risks of the estimators (3) and (5) are compared with the expected risk of the Bayes estimator and with the expected risk of the ML estimator. This comparison is performed for the Poisson distribution firstly with a Gamma distribution as prior and secondly with a one point distribution as prior. Six different numbers of past data are considered at the comparison, they are $n = 15, 30, 60, 120, 240, 480$.

Johns points out that the past samples can be of the form $\underline{z} = (z_{i1}, \dots, z_{i, p+k_i})$ where $k_i \geq 1$, $i = 1, \dots, n$. This means, that the number of observations of the

samples can vary. The additional information available when $k_i > 1$ can be used by substituting

$$u(z_{ip}) J(\underline{z}_i^{(p)} = \underline{z}^{(p)}) \quad (11)$$

for

$$J(\underline{z}_i^{(p)} = \underline{z}^{(p)}) (w_i/k_i) \sum_{j=p}^{p+k_i} u(z_{ij}) \quad (12)$$

and

$$\sum_{i=1}^n J(\underline{z}_i^{(p)} = \underline{z}^{(p)}) \quad (13)$$

for

$$\sum_{i=1}^n w_i J(\underline{z}_i^{(p)} = \underline{z}^{(p)}) . \quad (14)$$

The quantities $w_1, \dots, w_n$ are positive numerical weights depending on the k_i 's. It can be shown, the author writes, that whenever the w_i 's are uniformly bounded above and below by positive constants, the expected risk of the resulting estimator approaches the risk of the Bayes estimator as n becomes large. However, in general, when the k_i 's are not all the same, the optimum choice of $w_1, \dots, w_n$ depends on the unknown a priori probability structure and hence cannot be determined. The author does not furthermore treat the situation above and with the limited information, it is hard to use the proposed method in a concrete situation.

Krutchkoff (1967) has used Johns' procedure in a special case, which the author calls "supplementary sample EB approach". After the current observation z is obtained and the estimate of the current parameter is constructed we get supplementary information y concerning the current parameter. The distributions of z and y do not have to be the same, but it is stipulated that for

given θ , z and y are independent. Besides it is known that $E(y|\theta) = \theta$ and $E(y^2) < \infty$. The author presents the following equality

$$E(\theta|z) = E(y|z) . \quad (15)$$

Krutchkoff constructs the non-parametric EB estimator

$$\lambda(z) = \begin{cases} \frac{\sum_{i=1}^n y_i I(z_i = z)}{\sum_{i=1}^n I(z_i = z)} & \text{if } \sum_{i=1}^n I(z_i = z) > 0 \\ 0 & \text{otherwise} \end{cases} \quad (16)$$

where

$$I(D) = \begin{cases} 1 & \text{if } D \\ 0 & \text{otherwise} \end{cases}$$

and proves that it is asymptotically optimal.

Johns' estimator and Krutchkoff's type of Johns' estimator are not smooth. By using a regression model Krutchkoff is able to construct a smooth non-parametric EB estimator.

If $E(\theta|z) = \alpha + \beta z$, we get from (15) that $E(y|z) = \alpha + \beta z$. The parameters α and β are estimated from past data with the method of least squares and a non-parametric EB estimator is obtained. This is shown to be asymptotically optimal.

In the article, a Monte Carlo simulation is reported, where the parameter, θ , is a normal distributed variable with mean 0.2 and standard deviation 0.05. The variable, z , is binomial distributed with the parameters $(5, \theta)$. The supplementary information, y , is a binomial distributed variable with the parameters $(10, \theta)$. The mean squared error of the ML estimator, $z/5$, and the Bayes estimator are compared with the estimated mean squared error of the two non-parametric EB estimators proposed in the article. In the investigated situation, the mean squared errors of the two EB estimators are smaller than the mean squared error of the ML estimator. Even though the linearity assumption does not hold here, the regression estimator has a smaller mean squared error than does the estimator (16). This, the author writes, may suggest a practical

compromise. If the data look fairly linear one should use the regression estimator until a sufficient number of experiments have occurred, and then switch over to the estimator (16). This would ensure asymptotic optimality even when the linearity assumption does not hold.

Gabbert and Krutchkoff (1970) have used Krutchkoff's regression method to obtain control limits in a quality control situation. The authors reckon, as Krutchkoff does, on obtaining supplementary information, y , where $E(y|\theta) = \theta$, after the current data, z , have been obtained. They only use the last fifteen samples to constitute the estimates.

In the article, the authors mention a comparison they have made between the non-parametric EB procedure and the classical procedure. The problem is to determine if the production process is 'out of control' or not. The comparison is carried out with the aid of Monte Carlo simulations. The authors conclude that for small sample sizes and small to moderate departures from control, the supplementary sample non-parametric EB procedure will detect an out-of-control situation sooner than the usual control chart procedure.

The authors point out that if the number of past samples entering into the EB estimator is decreased the stopping time does not decrease. An increase of the number of past samples entering into the EB estimator, however, increases the stopping time. This surprising result is not enlarged upon further by the authors, but we have to think of the simulation situation which says that in the first fifteen samples the parameter is in-control but from the sixteenth sample the parameter stays out-of-control. This must imply that if all past samples are included in the calculation of the EB estimator the past samples in-control have an effect of the estimate so that the tendency to be in-control is strengthened. This effect is of course diminished quicker if only a certain number of the last samples is taken as past data in the EB estimator.

Ergle and Krutchkoff (1972) have proposed a non-parametric EB estimator which is based upon the above mentioned procedure of Lemon and Krutchkoff (1969), where the prior distribution is estimated by its empirical distribution function. In order to obtain a non-parametric estimator, the next step is to estimate the conditional function $f(z|\theta)$ with the aid of the already mentioned density estimator developed by Parzen (1962). If it will be possible to realize the estimation a sample of more than one observation for every θ -value has

to be present. Thus, with the notation used above we have $p > 1$. The number of observations in the samples has to be greater than one for every non-parametric procedure, with the exception of Krutchkoff's method where supplementary information is instead available.

The proposed procedure of Ergle and Krutchkoff derives from the Bayes estimator

$$\delta(z) = \frac{\int \theta f(z|\theta) dG(\theta)}{\int f(z|\theta) dG(\theta)} \quad (17)$$

If for the moment we suppose that the sequence θ_i , $i = 0, 1, \dots, n$, is known, the Bayes estimator can be approximated by

$$\pi(z) = \frac{\sum_{i=0}^n \theta_i f(z|\theta_i)}{\sum_{i=0}^n f(z|\theta_i)} \quad (18)$$

if $G(\theta)$ is estimated by its empirical distribution function. Of course, the θ_i , $i = 0, 1, \dots, n$ are not known, the authors write, but one usually has estimates of them. Call these estimates $\hat{\theta}_i$, $i = 0, 1, \dots, n$, and use them to estimate θ_i , $i = 0, 1, \dots, n$. Then (18) becomes

$$\pi(z) = \frac{\sum_{i=0}^n \hat{\theta}_i f(z|\hat{\theta}_i)}{\sum_{i=0}^n f(z|\hat{\theta}_i)} \quad (19)$$

which is the estimator proposed by Lemon and Krutchkoff, without iterating. In the situation dealt with by the authors, the $f(z|\hat{\theta}_i)$ for $i = 0, 1, \dots, n$, are unknown and without loss of generality can therefore be written in the more precise form $f(z|\theta_i)$, $i = 0, 1, \dots, n$. Here, however, z is a random sample, $\underline{z}$, and thus we can write

$$f(\underline{z}|\theta) = \prod_{j=1}^p f(z_j|\theta) . \quad (20)$$

In order to obtain the density estimator of Parzen, the following quantities are introduced

$$h_i = \left(\frac{1}{p_i}\right)^{7/10} \left\{ \sum_{j=1}^{p_i} (z_{ij} - \bar{z}_{i\cdot})^2 \right\}^{1/2} \quad (21)$$

where

$$\bar{z}_{i\cdot} = \frac{1}{p_i} \sum_{j=1}^{p_i} z_{ij} . \quad (22)$$

If

$$A_{ij}(z_k) = (z_k - z_{ij})/(2h_i) \quad (23)$$

then the proposed estimator for $f(z|\theta_i)$ is given by

$$\hat{f}(z_k|\theta_i) = 1/(2\pi p_i^{4/5}) \sum_{j=1}^{p_i} \left\{ \frac{\sin\{A_{ij}(z_k)\}}{A_{ij}(z_k)} \right\}^2 \quad (24)$$

with $\sin 0/0 = 1$. Using (20), we can now obtain the required EB estimator

$$\lambda(\underline{z}) = \frac{\sum_{i=0}^n \hat{\theta}_i \hat{f}(\underline{z}|\theta_i)}{\sum_{i=0}^n \hat{f}(\underline{z}|\theta_i)} \quad (25)$$

where

$$\hat{f}(\underline{z}|\theta_i) = \prod_{j=1}^p \hat{f}(z_j|\theta_i) \quad (26)$$

while p and $\underline{z}$ refer to the current sample.

The proposed procedure has an advantage over the non-parametric procedures presented in Chapter 3, in that the number of observations in the samples is variable. This is indicated by the notation p_i for the size of the i th sample. The authors note that the characteristics of θ are not specified. It can be a mean, a mode, a standard deviation, a percentile, or any other parameter for which there is an estimate. It ought to be pointed out here that this is true even for the procedures to be presented in Chapter 3. For estimating each of these parameters the only change in (25) will be the value of $\hat{\theta}_i$, $i = 0, 1, \dots, n$. The estimates $\hat{f}(\underline{z}|\theta_i)$ remain unchanged.

In the article, the authors mention that Monte Carlo simulations have been performed in order to compare the various classical estimators against the EB estimators for the mean, the variance, the standard deviation, the mode and the 80th percentile. The number of observations in the samples have varied from two to fifteen. The authors state, that the ratio of the mean squared error of the EB estimator and the particular classical estimator increases when p increases although the ratio generally remains less than one. This is due to the fact the EB procedures derive their excellent small-sample properties as a result of their ability to use more information to form a particular estimate than does the classical estimator. There was definite improvement in using EB estimators instead of classical estimators, but greater variability was noted when the number of observations in the past samples were allowed to vary randomly from two to fifteen.

Ergle and Krutchkoff write, "a nonparametric empirical Bayes procedure was suggested by Johns (1957) and another by Krutchkoff (1967). Johns requires a random sample of length p , as in this procedure but obtains an estimate of the posterior expectation based on a vector of length $p - 1$. Besides estimating a posterior expectation based on one observation less than available, the procedure has very poor small-sample, small number of past experiences, properties. Krutchkoff's supplementary sample nonparametric approach

assumes that there is some supplementary observation which becomes available, after each estimate is given, which can be used in subsequent estimations. Besides requiring a supplementary sample, Krutchkoff's procedure also has poor small-sample properties. On the other hand, the supplementary sample procedure can be shown to be asymptotically optimal. Both John's procedure and the one presented here are not asymptotically optimal in the usual sense but are asymptotically optimal when p is allowed to increase without bound.

The present procedure, however, unlike John's procedure, does have excellent small-sample properties even for vector lengths as small as two."

The following comment should be made in this connection. It is said that John's estimator is based only on $p-1$ observations. This is only partly true since the average (5) can be constructed. In Chapter 5 Monte Carlo simulations, which throw light upon the small-sample properties of different procedures, will be presented.

If we let the number of past samples increase, the expected risk of the EB estimator decreases substantially in the beginning if it is compared with the expected risk of the classical estimator, cf. the figures in Chapter 2.3. Ergle and Krutchkoff point out in the article that this decrease was noted to level off at about fifteen past experiences. For this reason, it is recommended that if trending is suspected one should use only the last fifteen sets of data in forming the EB estimate.

2.3 Design of Monte Carlo simulations of EB procedures

In order to get an idea of the inaccuracy of the EB estimators, Monte Carlo simulations are used so that the expected risks of the estimators can be estimated.

If the EB estimator in question is denoted by $\pi(z)$ the expected risk in the binomial case say, can be written as

$$E_n W \{ \pi(z) \} = \sum_{z_1=0}^p \dots \sum_{z_n=0}^p \left\{ \int \sum_{z=0}^p \{ \pi(z) - \theta \}^2 f(z | \theta) dG(\theta) \right\} \cdot f(z_n | \theta) \dots f(z_1 | \theta) \quad (1)$$

where the different $f(z|\theta)$ are binomial probabilities. In experiment $n + 1$ there are as many estimators, $\pi(z_0)$, as there are sequences $z_0, z_1, \dots, z_n$. The expected risk (1) is thus the risk of $\pi(z_0)$ averaged over the total number of possible sequences.

To compare the actual EB estimator with the appropriate classical estimator, one normally forms the ratio

$$\frac{\text{average squared error of the EB estimator}}{\text{average squared error of the classical estimator}} \quad (2)$$

This ratio is estimated with the aid of Monte Carlo simulations.

In the binomial case the ratio is expressed as

$$R_{n+1} = \frac{\sum_{j=1}^k \{\pi(z_{0j}) - \theta_{0j}\}^2 / k}{\sum_{j=1}^k \sum_{i=0}^n \{z_{ij}/p - \theta_{ij}\}^2 / (k(n+1))} \quad (3)$$

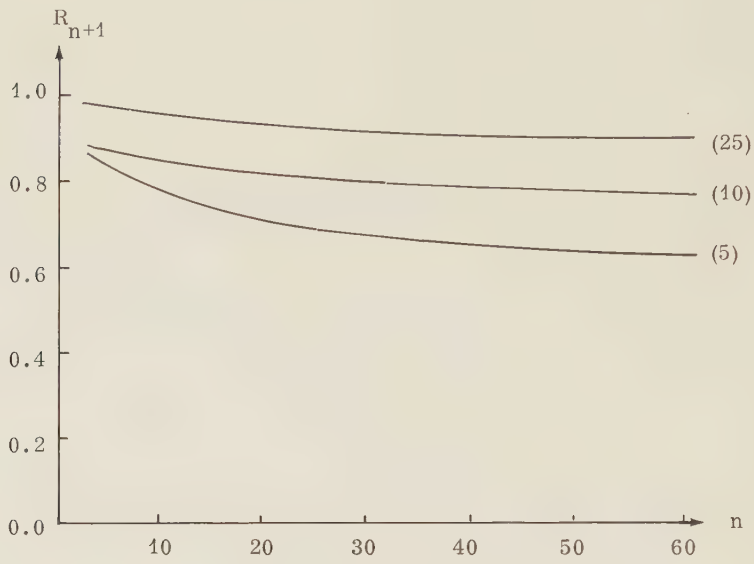
where $n + 1$ is the number of experiments and k is the number of iterations, i.e. the number of sequences of $n + 1$ experiments. Some authors do not estimate the denominator of the ratio (2) but form the ratio of the estimated expected risk of the EB estimator and the theoretical expected risk of the classical estimator.

In the Monte Carlo simulations, the prior distributions are often different Pearson curves, for example bell, J-shaped, L-shaped and U-shaped distributions. Almost every author states that the ratio (3) is influenced by the prior distribution almost entirely through the S factor. The S factor is defined as the ratio of the Bayes risk of the classical estimator, $\hat{\theta}$, to the variance of the parameter, θ ,

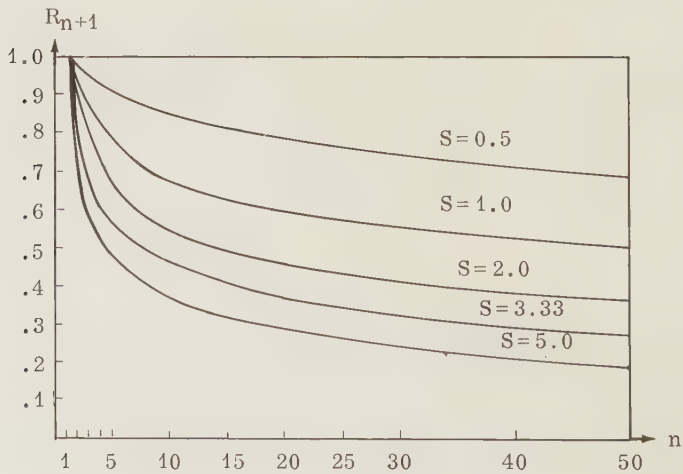
$$S = \frac{E\{\text{Var}(\hat{\theta}|\theta)\}}{\text{Var}(\theta)} \quad (4)$$

The authors have often chosen to illustrate graphically how the ratio, R_{n+1} , decreases when the number of past samples, n , increases. In Griffin and

Krutchkoff (1971 b) we find the following figure, where the ratio (3) is calculated for a special prior distribution and with the sample sizes $p = 5, 10, 25$.



Since the EB estimators are influenced by the factor S it will often be varied in the simulations and in the illustration figures. The figure below is from Lemon and Krutchkoff (1969).



3. NON-PARAMETRIC EB ESTIMATORS

EB procedures do not require that the prior distribution, $G(\theta)$, is known for the estimators to be constructed. However, the conditional distribution, $F(z|\theta)$, needs to be known. In non-parametric EB procedures, on the other hand, neither $G(\theta)$ nor $F(z|\theta)$ are required to be known. In order to estimate both $G(\theta)$ and $F(z|\theta)$ in a non-parametric EB situation we must obtain, for every realization of the parameter θ , more than one observation. These observations are divided into two parts. One part is used to construct the Bayes estimator and the other to estimate it.

3.1 Basic concepts

Past data are given by the vectors

$$\underline{z}_i = \begin{bmatrix} z_{i1} \\ \vdots \\ z_{ip} \end{bmatrix} \quad p > 1; i = 1, \dots, n; \quad Z = \begin{bmatrix} z'_1 \\ \vdots \\ z'_n \end{bmatrix} . \quad (1)$$

$Z^{(j)}$ denotes the matrix which remains when the row j has been omitted from the matrix Z .

Current data are given by the vector

$$\underline{z}_0 = \begin{bmatrix} z_{01} \\ \vdots \\ z_{0p} \end{bmatrix} \quad p > 1 . \quad (2)$$

$\underline{z}^{(j)}$ denotes the vector which remains when the element j has been omitted from the vector $\underline{z}$.

Past parameter values

$$\theta_i \quad i = 1, \dots, n \text{ are all unknown.} \quad (3)$$

The current parameter value

$$\theta_0 \quad \text{is unknown and will be estimated.} \quad (4)$$

Data and parameter values can be put together in the following way

$$\begin{array}{c|cccc} z_{01} & z_{11} & z_{21} & \cdots & z_{n1} \\ \vdots & \vdots & \vdots & & \vdots \\ z_{0p} & z_{1p} & z_{2p} & \cdots & z_{np} \\ \hline \theta_0 & \theta_1 & \theta_2 & \cdots & \theta_n \end{array} \quad (5)$$

The parameter values are generated from the prior distribution, $G(\theta)$. The observations $z_{ij} | \theta_i$ for $j = 1, \dots, p$ are independently generated from the distribution $F(z | \theta_i)$ for $i = 0, 1, \dots, n$.

The observations z_i for $i = 0, 1, \dots, n$ are independently generated from the distribution

$$F(\underline{z}) = \int_{\theta} F(\underline{z} | \theta) dG(\theta) . \quad (6)$$

If f denotes the probability function of the distribution F , the simultaneous probability function of $(\underline{z}_0, \underline{z}_1, \dots, \underline{z}_n)$ can be expressed as

$$f(\underline{z}_0) \cdots f(\underline{z}_n) \quad (7)$$

where

$$f(\underline{z}_i) = \int_{\theta_i} f(z_{i1} | \theta_i) \cdots f(z_{ip} | \theta_i) dG(\theta_i) \quad i = 0, 1, \dots, n . \quad (8)$$

3.2 Direct estimation of the Bayes estimator

The basic estimation procedure in this Chapter comes from Johns (1957). The author has proved, as mentioned in Chapter 2.2, that the estimator (3.2.1 (27)) is asymptotically optimal both for discrete and continuous conditional distribution, $F(z|\theta)$. In this work, however, only discrete conditional distributions will be used. The notation, $f(z|\theta)$, will from now on be reserved for discrete probabilities.

3.2.1 γ procedure

The simplest case, where the non-parametric EB estimation can be applied, is when the past data consist of pairs of observations, i.e. when $p = 2$. The current data consist, however, of just one observation.

Past data are given by

$$\underline{z}_i = \begin{bmatrix} z_{i1} \\ z_{i2} \end{bmatrix}, \quad i = 1, \dots, n; \quad Z = \begin{bmatrix} z'_1 \\ \vdots \\ z'_n \end{bmatrix}. \quad (1)$$

In this case $Z^{(j)}$ is a row vector cf. Chapter 3.1.

Current data consist of only one observation

$$\underline{z}_0^{(2)} = z_0. \quad (2)$$

Data and parameter values can be put together in the tableau

$$\begin{array}{c|cccc} z_{01} = z_0 & z_{11} & z_{21} & \cdots & z_{n1} \\ & z_{12} & z_{22} & \cdots & z_{n2} \\ \theta_0 & \theta_1 & \theta_2 & \cdots & \theta_n \end{array} \quad (3)$$

Every pair of past observations can be looked upon, for a certain θ_i , as realizations of two independent stochastic variables z_1 and z_2 each of which

has the distribution function $F(z|\theta_i)$. Observe that it is true for every i that the variables z_1 and z_2 , unconditional of θ , do not require to be independent.

Suppose the function u has the qualities

$$E\{u(z)|\theta\} = \theta \quad (4)$$

and

$$E\{u^2(z)\} < \infty. \quad (5)$$

The second property is needed so that the estimator below will be asymptotically optimal.

By using the identity (4) and the conditional independence between the variables z_1 and z_2 the Bayes estimator can be expressed as the expected value of $u(z_2)$ given z_1 . Consequently we get

$$\begin{aligned} E(\theta|z_1) &= \int \theta g(\theta|z_1) d\theta = \int \theta \frac{f(z_1|\theta)}{f(z_1)} dG(\theta) = \\ &= \int \int u(z_2) f(z_2|\theta) \frac{f(z_1|\theta)}{f(z_1)} dz_2 dG(\theta) = \\ &= \int \int u(z_2) \frac{f(z_1, z_2|\theta)}{f(z_1)} dz_2 dG(\theta) = \int u(z_2) f(z_2|z_1) dz_2 = \\ &= E\{u(z_2)|z_1\}. \end{aligned} \quad (6)$$

Now the expected value, $E\{u(z_2)|z_0\}$, can be estimated with the corresponding sample moment and the following non-parametric EB estimator, $\gamma(z_0)$, can be defined as

$$\gamma(z_0) = \begin{cases} \frac{\sum_{i=1}^n u(z_{i2}) I(z_{i1} = z_0)}{m(Z^{(2)}, z_0)} & \text{if } m(Z^{(2)}, z_0) > 0 \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

where

$$I(D) = \begin{cases} 1 & \text{if } D \\ 0 & \text{otherwise} \end{cases}$$

and

$$m(Z^{(2)}, z_0) = \sum_{i=1}^n I(z_{i1} = z_0) .$$

The estimator (7) is asymptotically optimal. This can be shown with the aid of the extra risk. What will be shown is then, that the extra risk, $r(\gamma)$, of the estimator $\gamma(z_0)$ has the limit 0 when $n \rightarrow \infty$

$$\lim_{n \rightarrow \infty} r(\gamma) = \lim_{n \rightarrow \infty} E_{Z, z_0} \{ \gamma(z_0) - \delta(z_0) \}^2 = 0 \quad (8)$$

where $\delta(z_0)$ denotes the Bayes estimator, $E(\theta | z_0)$.

When $m(Z^{(2)}, z) = m > 0$, $\gamma(z)$ is simply the average of m independent unbiased estimates of $E\{u(z_2) | z\}$. According to (6), $E\{u(z_2) | z\}$ is equal to the Bayes estimator $\delta(z)$. Consequently

$$E_Z \{ \gamma(z) | m > 0 \} = \delta(z) . \quad (9)$$

If the difference

$$\gamma(z) - \delta(z) = \frac{\sum_i \{u(z_{i2}) - \delta(z)\} I(z_{i1} = z)}{\sum_i I(z_{i1} = z)} \quad (10)$$

is formed the following equalities can be established

$$\begin{aligned} E_Z \{\gamma(z) - \delta(z)\}^2 &= \\ &= E_{z_{11}, \dots, z_{n1}} \frac{\sum_{i,j} \text{Cov}\{u(z_{i2}), u(z_{j2}) | z_{i1}, z_{j1}\} I(z_{i1} = z_{j1} = z)}{(\sum_i I(z_{i1} = z))^2} = \\ &= E_{z_{11}, \dots, z_{n1}} \frac{\sum_i \text{Var}\{u(z_{i2}) | z_{i1}\} I(z_{i1} = z)}{(\sum_i I(z_{i1} = z))^2}. \end{aligned} \quad (11)$$

If m is allowed to take all possible values we get

$$\begin{aligned} E_Z \{\gamma(z) - \delta(z)\}^2 &= \\ &= \sum_{m=1}^n P(\sum_{i=1}^n I(z_{i1} = z) = m) \frac{\sum_i V(z_{i1}) I(z_{i1} = z)}{m^2} + \\ &+ P(\sum_i I(z_{i1} = z) = 0) \delta^2(z) = \\ &= \sum_{m=1}^n \binom{n}{m} p(z)^m q(z)^{n-m} \frac{V(z)}{m} + q(z)^n \delta^2(z) = \\ &= \frac{1}{p(z)} \sum_{m=1}^n \frac{m+1}{n+1} \binom{n+1}{m+1} p(z)^{m+1} q(z)^{n-m} \frac{V(z)}{m} + \\ &+ q(z)^n \delta^2(z) \leq \frac{2V(z)}{(n+1)p(z)} + q(z)^n \delta^2(z) \end{aligned} \quad (12)$$

because

$$\frac{m+1}{m} = 1 + \frac{1}{m} \leq 2 ,$$

and where

$$\begin{cases} V(z) = \text{Var}(u(z_2) | z) \\ p(z) = \sum_{z_2} f(z, z_2) \\ q(z) = 1 - p(z) \end{cases} . \quad (13)$$

If the estimator, $\gamma(z_0)$, is to be asymptotically optimal the expression (14) has to tend towards 0 as n increases.

$$\begin{aligned} \sum_{z_0} \left\{ \frac{2V(z_0)}{(n+1)p(z_0)} + q(z_0)^n \delta^2(z_0) \right\} p(z_0) &= \\ &= \sum_{z_0} \left\{ \frac{2V(z_0)}{n+1} + p(z_0) q(z_0)^n \delta^2(z_0) \right\} . \end{aligned} \quad (14)$$

The following inequalities can be formed

$$\begin{aligned} \frac{2V(z_0)}{n+1} + p(z_0) q(z_0)^n \delta^2(z_0) &\leq \frac{2V(z_0)}{n+1} + \delta^2(z_0) \leq V(z_0) + \delta^2(z_0) = \\ &= E\{u^2(z_2) | z_0\} . \end{aligned} \quad (15)$$

Since

$$E_{z_0} E_{z_2 | z_0} \{u^2(z_2) | z_0\} = E u^2(z_2) \quad (16)$$

and $E\{u^2(z_2)\}$ is finite due to the assumptions, the sum (14) will be dominated by a finite expression. It is also true that the terms in the sum (14) tend toward 0 as n increases. The limit of the whole sum (14) is thus 0 as $n \rightarrow \infty$. This means that the estimator, $\gamma(z_0)$, in (7) is asymptotically optimal.

Johns (1957) points out that the procedure probably is very ineffective when n is small. He suggests a modified procedure similar to that proposed below. When n is small then also $m(Z^{(2)}, z_0)$ will be small. This entails that the estimator, $\gamma(z_0)$, can be built on very few observations. In order to correct this, $\gamma(z_0)$ can be exchanged for an estimator of θ which depends only on the current observation z_0 when small values of $m(Z^{(2)}, z_0)$ are at hand. A modified non-parametric EB estimator, $\gamma^{(c)}(z_0)$, can be defined, for fixed $c \geq 0$, as

$$\gamma^{(c)}(z_0) = \begin{cases} \gamma(z_0) & \text{if } m(Z^{(2)}, z_0) > c \\ v(z_0) & \text{if } m(Z^{(2)}, z_0) \leq c \end{cases} \quad (17)$$

The modified estimator, $\gamma^{(c)}(z_0)$, is also asymptotically optimal. In order to prove this only small changes are required in the proof worked out above. If $c = 0$ the extra risk of the estimator (17) becomes

$$E_{z_0, Z} \{ \gamma^{(0)}(z_0) - \delta(z_0) \}^2 = \quad (18)$$

$$\begin{aligned} &= E_{z_0} \left\{ \sum_{m=1}^n \binom{n}{m} p(z_0)^m q(z_0)^{n-m} \frac{V(z_0)}{m} + q(z_0)^n \{u(z_0) - \delta(z_0)\}^2 \right\} \leq \\ &\leq \sum_{z_0} \left\{ \frac{2V(z_0)}{(n+1)p(z_0)} + q(z_0)^n \{u(z_0) - \delta(z_0)\}^2 \right\} p(z_0) . \end{aligned}$$

The sum is thus dominated by

$$\begin{aligned} E_{z_0} \{ V(z_0) + \{u(z_0) - \delta(z_0)\}^2 \} &= E_{z_0} \{ E(u^2(z_2) | z_0) + u^2(z_0) - \\ &- 2u(z_0) \delta(z_0) \} \leq E u^2(z_2) + E u^2(z_0) - 2 \sqrt{E u^2(z_0) E \delta^2(z_0)} \end{aligned} \quad (19)$$

where the last member of (19) has been obtained by the Schwartz inequality. If the expected value, $E\{\delta^2(z_0)\}$, exists the estimator $\gamma^{(0)}(z_0)$ is asymptotically optimal.

If $c = a > 0$ and a is an integer, the extra risk of the estimator (17) is equal to

$$\begin{aligned}
 E_{z_0, Z} \{ \gamma^{(a)}(z_0) - \delta(z_0) \}^2 &= E_{z_0} \left\{ \sum_{m=a+1}^n p(z_0)^m q(z_0)^{n-m} \frac{V(z_0)}{m} + \right. \\
 &+ (q(z_0)^n + p(z_0) q(z_0)^{n-1} + \dots + p(z_0)^a q(z_0)^{n-a}) \cdot \\
 &\cdot \{ u(z_0) - \delta(z_0) \}^2 \} \leq \sum_{z_0} \left\{ \frac{2V(z_0)}{(n+1)p(z_0)} + (q(z_0)^n + \right. \\
 &+ p(z_0) q(z_0)^{n-1} + \dots + p(z_0)^a q(z_0)^{n-a}) \{ u(z_0) - \delta(z_0) \}^2 \} p(z_0) .
 \end{aligned} \tag{20}$$

The sum is also now dominated by the expression (19) and the estimator, $\gamma^{(a)}(z_0)$, is asymptotically optimal if the expected value, $E\{\delta^2(z_0)\}$ exists.

The situation which has been described above, where the current data consist of one observation while the past data consist of n pairs of observations, is probably very unusual in reality. It is natural to think that the data consist of samples of p observations which are generated, one for every θ_i , for $i = 0, 1, \dots, n$. Thus, the number of observations in past and current data is the same. In the situation described above the current data consist of two observations z_{01} and z_{02} . The estimators (7) and (17), however, use only one of z_{01} and z_{02} . It is irrelevant if the current observation in (7) and (17) equals z_{01} or z_{02} . The only difference is that the index j in $Z^{(j)}$ is changed to 2 and 1 respectively in order to attain correspondance. That the estimators have this limitation, is not desirable since all the information is not used. Johns (1957) has proposed the average estimator (2.2(5)). We use his proposition and construct the estimators γ and $\gamma^{(c)}$ where both of z_{01} and z_{02} are included

$$\gamma = \frac{1}{2} \gamma(z_{01}) + \frac{1}{2} \gamma(z_{02}) \quad (21)$$

and

$$\gamma^{(c)} = \frac{1}{2} \gamma^{(c)}(z_{01}) + \frac{1}{2} \gamma^{(c)}(z_{02}) . \quad (22)$$

The estimators (21) and (22) can be shown to be asymptotically optimal if the Bayes estimator is defined as

$$\frac{1}{2} \delta(z_{01}) + \frac{1}{2} \delta(z_{02}) . \quad (23)$$

Now it is usually true that

$$\delta(\underline{z}) \neq \frac{1}{2} \delta(z_{01}) + \frac{1}{2} \delta(z_{02}) \quad (24)$$

but from Johns' inequalities (2.2(9) and (10)) we can conclude that

$$\mathbb{E}_Z W(\gamma) \leq \mathbb{E}_Z W(\gamma(z_1)) \quad (25)$$

and that

$$W\left(\frac{1}{2} \delta(z_1) + \frac{1}{2} \delta(z_2)\right) \leq W(\delta(z_1)) . \quad (26)$$

Thus, the estimators (21) and (22) should be preferred to (7) and (17), respectively.

The special case, where p equals 2, which has been treated, will now be extended in order to cover the general case where $p \geq 2$. Firstly an estimator corresponding to (7) will be constructed, in the case that the current data consist of only $p - 1$ observations. Secondly an estimator corresponding to (21) will be constructed, in the general case when the current data consist of p observations.

Let the current data be $\underline{z}_0^{(p)} = (z_{01}, \dots, z_{0,p-1})'$. When $p \geq 2$ the estimator (7) for $p = 2$ will correspond to

$$\gamma(\underline{z}_0^{(p)}) = \begin{cases} \frac{\sum_{i=1}^n u(z_{ip}) J(\underline{z}_i^{(p)} = \underline{z}_0^{(p)})}{m(\underline{Z}^{(p)}, \underline{z}_0^{(p)})} & \text{if } m(\underline{Z}^{(p)}, \underline{z}_0^{(p)}) > 0 \\ 0 & \text{otherwise} \end{cases} \quad (27)$$

where

$$J(\underline{z}_i^{(p)} = \underline{z}_0^{(p)}) = \begin{cases} 1 & \text{if any of the permutations of} \\ & (z_{i1}, \dots, z_{i,p-1})' \text{ equals} \\ & (z_{01}, \dots, z_{0,p-1})' \\ 0 & \text{otherwise} \end{cases} \quad (28)$$

and

$$m(\underline{Z}^{(p)}, \underline{z}_0^{(p)}) = \sum_{i=1}^n J(\underline{z}_i^{(p)} = \underline{z}_0^{(p)}) . \quad (29)$$

The estimator (27) is asymptotically optimal. This is shown in the same way as for the estimator (7). Certain quantities will of course be changed in the proof. The current observation z_0 is replaced by the vector $\underline{z}_0^{(p)}$. The observations z_{i1} are replaced by the vectors $\underline{z}_i^{(p)}$ and z_{i2} are replaced by z_{ip} where $i = 1, \dots, n$. The variance and the probabilities in the expression (13) now become

$$\begin{cases} V(\underline{z}^{(p)}) = \text{Var}(u(\underline{z}_p) | \underline{z}^{(p)}) \\ p(\underline{z}^{(p)}) = \sum_{\substack{z_p \\ i_1 \neq i_2 \neq \dots \neq i_{p-1}}} \sum f(z_{i_1}, z_{i_2}, \dots, z_{i_{p-1}}, z_p) \\ q(\underline{z}^{(p)}) = 1 - p(\underline{z}^{(p)}) \end{cases} \quad (30)$$

The proof can now be used and the estimator (27) is asymptotically optimal.

The indicator in (27) can be defined as

$$I(\underline{z}_i^{(p)} = \underline{z}_0^{(p)}) .$$

This means that no permutation of the elements in $\underline{z}_i^{(p)}$ is made. Since the observations in $\underline{z}_i^{(p)}$ are independent, it is not appropriate that the order of the observations should be able to determine whether $u(z_{ip})$ will be in the estimator or not.

Another way to define the indicator is to construct

$$\sum_{\underline{P}} I(\underline{P}\underline{z}_i^{(p)} = \underline{z}_0^{(p)})$$

where $\underline{P}\underline{z}_i^{(p)}$ means all permutations of the elements of the vector $\underline{z}_i^{(p)}$. This way of constructing the indicator implies that the quantity $u(z_{ip})$ will be summed up the number of times for which there is a permutation of the elements in $\underline{z}_i^{(p)}$ equal to $\underline{z}_0^{(p)}$.

In this thesis, the indicator (28) will be used, which is the same as that appearing in Johns (1957).

The modified estimator (17), which is suitable when n is small, is in the general case, when $p \geq 2$, equal to

$$\gamma^{(c)}(\underline{z}_0^{(p)}) = \begin{cases} \gamma(\underline{z}_0^{(p)}) & \text{if } m(Z^{(p)}, \underline{z}_0^{(p)}) > c \\ v(\underline{z}_0^{(p)}) & \text{if } m(Z^{(p)}, \underline{z}_0^{(p)}) \leq c \end{cases} \quad (31)$$

where $v(\underline{z}_0^{(p)})$ denotes a conventional estimator of θ , which is built only out of the current data. The estimator (31) can be shown to be asymptotically optimal. This proof is carried out the same way as that for the estimator (17), provided that we make the same modifications, as we did above when asymptotic optimality of the estimator (27) was to be shown.

If the past data are built up in such a way that p observations are generated

for every realization of θ , it is reasonable to assume that the current data also consist of p observations. Consequently, the situation is contrary to the above, where the sample of the current data consist of one observation less than the samples of the past data. When the past and the current data consist of p -vectors, it is possible to construct p different $Z^{(j)}$ and $\underline{z}_0^{(j)}$ for $j = 1, \dots, p$. It is natural to combine all the p estimators which can be constructed in the average

$$\gamma_p = \frac{1}{p} \sum_{j=1}^p \gamma(\underline{z}_0^{(j)}) \quad (32)$$

and

$$\gamma_p^{(c)} = \frac{1}{p} \sum_{j=1}^p \gamma^{(c)}(\underline{z}_0^{(j)}) . \quad (33)$$

The estimators (32) and (33) should be preferred to (27) and (31), respectively. The reasons for this statement were given in the connection with the estimators (21) and (22). These reasons can easily be extended to hold for p -vectors.

The partition of the samples into two parts, where the parts are composed of $p - 1$ and 1 observations, which has been carried out above, is proposed by Johns (1957). The partition can be made more general. The following notation is introduced.

$$\Omega = \{1, \dots, p\}$$

$$A \subset \Omega$$

The notation $|A| = r$ means that the number of elements in A equals r , where $1 \leq r \leq p - 1$. (34)

For every r there are $\binom{p}{r}$ different A and totally there

$$\text{are } \sum_{r=1}^{p-1} \binom{p}{r} = 2^p - 2 \text{ different } A.$$

Thus

$$A \neq \{ \} \& \{1, \dots, p\} .$$

The complement of A is denoted by $\overline{A}$.

Let, for a given A , the function u be such that

$$E\{u(\underline{z}^{(A)})|\theta\} = \theta \quad (35)$$

and

$$E\{u^2(\underline{z}^{(A)})|\theta\} < \infty \quad (36)$$

where the notation, $\underline{z}^{(A)}$, as above, denotes the vector which remains, when the r variables, whose indices are in A , have been omitted from the vector $\underline{z}$. The non-parametric EB estimators become in this case

$$\gamma(\underline{z}_0^{(\bar{A})}) = \begin{cases} \frac{\sum_{i=1}^n u(\underline{z}_i^{(A)}) J(\underline{z}_i^{(\bar{A})} = \underline{z}_0^{(\bar{A})})}{m(\underline{z}^{(\bar{A})}, \underline{z}_0^{(\bar{A})})} & \text{if } m(\underline{z}^{(\bar{A})}, \underline{z}_0^{(\bar{A})}) > 0 \\ 0 & \text{otherwise} \end{cases} \quad (37)$$

and

$$\gamma^{(c)}(\underline{z}_0^{(\bar{A})}) = \begin{cases} \gamma(\underline{z}_0^{(\bar{A})}) & \text{if } m(\underline{z}^{(\bar{A})}, \underline{z}_0^{(\bar{A})}) > c \\ v(\underline{z}_0) & \text{if } m(\underline{z}^{(\bar{A})}, \underline{z}_0^{(\bar{A})}) \leq c \end{cases} \quad (38)$$

The estimators (37) and (38) are asymptotically optimal if the Bayes estimator is defined as $\delta(\underline{z}_0^{(\bar{A})})$. This can easily be shown with the aid of the results gained above, since the estimator (27) is a special case of the estimator (37) and the estimator (31) is a special case of the estimator (38). To let the Bayes estimator be built up only out of $\underline{z}_0^{(\bar{A})}$ is unnatural, since the current data consist of a p -vector $\underline{z}_0$. The Bayes estimator should then be defined as $\delta(\underline{z}_0)$. Thus, the estimators are not asymptotically optimal in the usual sense.

To every r , $\binom{p}{r}$ different A can be constructed. Thus, $\binom{p}{r}$ different estimators of the type (37) and (38) can be constructed. These $\binom{p}{r}$ estimators can

be combined in a natural way by an average. There is no reason to prefer some, since the partitions, for a given r , are symmetric.

The two average estimators become

$$\gamma_r = \frac{1}{\binom{p}{r}} \sum_{\substack{A \\ |A|=r}} \gamma(z_0^{(\bar{A})}) \quad (39)$$

and

$$\gamma_r^{(c)} = \frac{1}{\binom{p}{r}} \sum_{\substack{A \\ |A|=r}} \gamma^{(c)}(z_0^{(\bar{A})}) . \quad (40)$$

With the aid of the above results for the estimators (32) and (33) it can be shown that the estimators (39) and (40) should be preferred to the estimators (37) and (38) respectively. Since r can take the values $1, \dots, p-1$, $p-1$ different estimators of the type (39) and (40) can be constructed. The question arises if some r -value is optimal, which means that the corresponding estimator has the least expected risk among the $p-1$ estimators, or if there is an optimal combination of the $p-1$ estimators.

3.2.2 ϵ procedure

It is possible to modify the procedure, which has been treated in Chapter 3.2.1, by permuting the past data. This attempt has been proposed by Maritz (1970). We first consider the simplest case of data, when non-parametric EB estimators can be used. It is described in the tableau (3.2.1(3)). Since past observations are randomly generated for every parameter value, there is no reason for matching z_0 with $Z^{(2)}$ rather than $Z^{(1)}$. The current observation z_0 can be matched with both of them. The non-parametric EB estimator which corresponds to the estimator (3.2.1(7)) then becomes

$$\varepsilon(z_0) = \begin{cases} \frac{\sum_{i=1}^n \{u(z_{i2}) I(z_{i1} = z_0) + u(z_{i1}) I(z_{i2} = z_0)\}}{m(Z^{(2)}, z_0) + m(Z^{(1)}, z_0)} & \text{if } m(Z^{(2)}, z_0) + \\ & + m(Z^{(1)}, z_0) > 0 \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

The expression above can be written as a linear combination of γ estimators.

We get

$$\varepsilon(z_0) = \begin{cases} \frac{m(Z^{(2)}, z_0) \gamma(Z^{(2)}, z_0) + m(Z^{(1)}, z_0) \gamma(Z^{(1)}, z_0)}{m(Z^{(2)}, z_0) + m(Z^{(1)}, z_0)} & \text{if } m(Z^{(2)}, z_0) + \\ & + m(Z^{(1)}, z_0) > 0 \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

Here, $\gamma(Z^{(2)}, z_0)$, stands for the expression (3.2.1(7)), where z_0 is matched with $Z^{(2)}$, while z_0 is matched with $Z^{(1)}$ in $\gamma(Z^{(1)}, z_0)$. Thus

$$\gamma(Z^{(2)}, z_0) = \frac{\sum_{i=1}^n u(z_{i2}) I(z_{i1} = z_0)}{m(Z^{(2)}, z_0)} \quad (3)$$

and

$$\gamma(Z^{(1)}, z_0) = \frac{\sum_{i=1}^n u(z_{i1}) I(z_{i2} = z_0)}{m(Z^{(1)}, z_0)} \quad (4)$$

where

$$m(Z^{(2)}, z_0) = \sum_{i=1}^n I(z_{i1} = z_0) \quad (5)$$

and

$$m(Z^{(1)}, z_0) = \sum_{i=1}^n I(z_{i2} = z_0) . \quad (6)$$

Because the estimator, $\epsilon(z_0)$, can be written as a linear combination of the asymptotically optimal estimators treated above, we can establish that the estimator, $\epsilon(z_0)$, also is asymptotically optimal.

When the number of past samples, n , is small, the same modification as above can be made

$$\epsilon^{(c)}(z_0) = \begin{cases} \epsilon(z_0) & \text{if } m(Z^{(2)}, z_0) + m(Z^{(1)}, z_0) > c \\ v(z_0) & \text{if } m(Z^{(2)}, z_0) + m(Z^{(1)}, z_0) \leq c \end{cases} \quad (7)$$

In order to prove the asymptotic optimality of the estimator (7) we use the possibility of writing it as a linear combination of the asymptotically optimal estimators treated above. Observe that $\gamma(Z^{(2)}, z_0)$ is changed to $v(z_0)$ if $m(Z^{(2)}, z_0) \leq c - m(Z^{(1)}, z_0)$ and not if $m(Z^{(2)}, z_0) \leq c$, the corresponding thing is valid for $\gamma(Z^{(1)}, z_0)$.

In the more general case, when the past and the current data consist of p -vectors, the average estimator, ϵ , is constructed, as for the γ -procedure when $p = 2$, as

$$\epsilon = \frac{1}{2} \epsilon(z_{01}) + \frac{1}{2} \epsilon(z_{02}) \quad (8)$$

where

$$\epsilon(z_{01}) = \begin{cases} \frac{m(Z^{(2)}, z_{01}) \gamma(Z^{(2)}, z_{01}) + m(Z^{(1)}, z_{01}) \gamma(Z^{(1)}, z_{01})}{m(Z^{(2)}, z_{01}) + m(Z^{(1)}, z_{01})} & \text{if} \\ & m(Z^{(2)}, z_{01}) + \\ & + m(Z^{(1)}, z_{01}) > 0 \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

and

$$\epsilon(z_{02}) = \begin{cases} \frac{m(Z^{(2)}, z_{02}) \gamma(Z^{(2)}, z_{02}) + m(Z^{(1)}, z_{02}) \gamma(Z^{(1)}, z_{02})}{m(Z^{(2)}, z_{02}) + m(Z^{(1)}, z_{02})} & \text{if} \\ & m(Z^{(2)}, z_{02}) + \\ & + m(Z^{(1)}, z_{02}) > 0 \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

The explanation for constructing the average in (8), is analogous to that given for the corresponding average of the γ procedure in (3.2.1(21)).

The extensions for the γ procedure in (3.2.1(27), (31), (37), (38)) can of course be carried out also for the ϵ procedure. All these ϵ estimators are asymptotically optimal in the same sense as before. This is easily shown since the ϵ estimators can be expressed as linear combinations of γ estimators. The average estimators which correspond to those shown in (3.2.1(32), (33), (39), (40)) can be formed. That these average estimators are appropriate also for the ϵ procedure is justified in the same way as for the γ procedure.

Let B have the same meaning as A (3.2.1(34)). The estimator $\epsilon(\underline{z}_0^{(\bar{A})})$, which corresponds to the estimator, $\gamma(\underline{z}_0^{(\bar{A})})$ in (3.2.1(37)), can then be expressed as

$$\epsilon(\underline{z}_0^{(\bar{A})}) = \begin{cases} \frac{\sum_{\substack{B \\ |B|=r}} k_B \gamma(Z^{(\bar{B})}, \underline{z}_0^{(\bar{A})})}{\sum_{\substack{B \\ |B|=r}} (Z^{(\bar{B})}, \underline{z}_0^{(\bar{A})})} & \text{if} \\ & \sum_{\substack{B \\ |B|=r}} (Z^{(\bar{B})}, \underline{z}_0^{(\bar{A})}) > 0 \\ 0 & \text{otherwise} \end{cases} \quad (11)$$

where

$$k_B = \frac{m(Z^{(\bar{B})}, \underline{z}_0^{(\bar{A})})}{\sum_{\substack{B \\ |B|=r}} m(Z^{(\bar{B})}, \underline{z}_0^{(\bar{A})})} \quad (12)$$

and where

$$\gamma(Z^{(\bar{B})}, \underline{z}_0^{(\bar{A})}) = \frac{\sum_{i=1}^n u(\underline{z}_i^{(B)}) J(\underline{z}_i^{(\bar{B})} = \underline{z}_0^{(\bar{A})})}{m(Z^{(\bar{B})}, \underline{z}_0^{(\bar{A})})} \quad (13)$$

and

$$m(Z^{(\bar{B})}, \underline{z}_0^{(\bar{A})}) = \sum_{i=1}^n J(\underline{z}_i^{(\bar{B})} = \underline{z}_0^{(\bar{A})}) . \quad (14)$$

Corresponding to the average estimator, γ_r , in (3.2.1(39)) we have

$$\epsilon_r = \frac{1}{\binom{p}{r}} \sum_{\substack{A \\ |A|=r}} \epsilon(\underline{z}_0^{(\bar{A})}) . \quad (15)$$

If the expressions (7) - (14) are not used we get

$$\epsilon_r = \begin{cases} \frac{1}{\binom{p}{r}} \sum_{\substack{A \\ |A|=r}} \frac{\sum_{\substack{B \\ |B|=r}} \sum_{i=1}^n u(\underline{z}_i^{(B)}) J(\underline{z}_i^{(\bar{B})} = \underline{z}_0^{(\bar{A})})}{\sum_{\substack{B \\ |B|=r}} \sum_{i=1}^n J(\underline{z}_i^{(\bar{B})} = \underline{z}_0^{(\bar{A})})} & \text{if} \\ \sum_{\substack{A \\ |A|=r}} \sum_{\substack{B \\ |B|=r}} \sum_{i=1}^n J(\underline{z}_i^{(\bar{B})} = \underline{z}_0^{(\bar{A})}) > 0 \\ 0 & \text{otherwise} \end{cases} \quad (16)$$

Corresponding to the modified average estimator, $\gamma_r^{(c)}$, in (3.2.1(40)) we have

$$\epsilon_r^{(c)} = \begin{cases} \epsilon_r & \text{if } \sum_{\substack{A \\ |A|=r}} \sum_{\substack{B \\ |B|=r}} \sum_{i=1}^n J(\underline{z}_i^{(\bar{B})} = \underline{z}_0^{(\bar{A})}) > c \\ v(\underline{z}_0) & \text{if } \sum_{\substack{A \\ |A|=r}} \sum_{\substack{B \\ |B|=r}} \sum_{i=1}^n J(\underline{z}_i^{(\bar{B})} = \underline{z}_0^{(\bar{A})}) \leq c \end{cases} \quad (17)$$

3.3 An example

In order to illustrate how one calculates some of the estimators, presented in Chapters 3.2.1 and 3.2.2, we give the following example. Suppose that θ , the proportion of pages affected by one or more misprints in this thesis, is to be estimated. The printer states, that in the four last theses which have been printed, three pages have been randomly drawn from each thesis. The pages drawn are then examined for misprints. The results of this check are as follows $\underline{z}_1 = (0, 0, 0)'$, $\underline{z}_2 = (0, 1, 0)'$, $\underline{z}_3 = (1, 0, 0)'$ and $\underline{z}_4 = (1, 1, 1)'$, where 1 means that the page drawn has one or more misprints while 0 means no misprint.

By drawing three pages randomly in this thesis and by examining the pages drawn, for misprints, it is possible to estimate θ with an EB procedure. Let the current data be $\underline{z}_0 = (1, 0, 0)'$. Suppose that the observations of each sample are independent.

As an illustration, the following two estimators from Chapter 3.2.1 are calculated, namely (37)

$$\gamma(\underline{z}_0^{(\bar{A})}) = \begin{cases} \frac{\sum_{i=1}^n u(\underline{z}_i^{(A)}) J(\underline{z}_i^{(\bar{A})} = \underline{z}_0^{(\bar{A})})}{\sum_{i=1}^n J(\underline{z}_i^{(\bar{A})} = \underline{z}_0^{(\bar{A})})} & \text{if } \sum_{i=1}^n J(\underline{z}_i^{(\bar{A})} = \underline{z}_0^{(\bar{A})}) > 0 \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

and the average estimator (39)

$$\gamma_r = \frac{1}{\binom{p}{r}} \sum_{\substack{A \\ |A|=r}} \gamma(\underline{z}_0^{(\bar{A})}) . \quad (2)$$

From Chapter 3.2.2 the corresponding two estimators are calculated, namely (11)

$$\epsilon(\underline{z}_0^{(\bar{A})}) = \left\{ \begin{array}{ll} \frac{\sum_{\substack{B \\ |B|=r}} \sum_{i=1}^n u(\underline{z}_i^{(B)}) J(\underline{z}_i^{(\bar{B})} = \underline{z}_0^{(\bar{A})})}{\sum_{\substack{B \\ |B|=r}} \sum_{i=1}^n J(\underline{z}_i^{(\bar{B})} = \underline{z}_0^{(\bar{A})})} & \text{if} \\ \sum_{\substack{B \\ |B|=r}} \sum_{i=1}^n J(\underline{z}_i^{(\bar{B})} = \underline{z}_0^{(\bar{A})}) > 0 \\ 0 & \text{otherwise} \end{array} \right. \quad (3)$$

and the average estimator (16)

$$\epsilon_r = \frac{1}{\binom{p}{r}} \sum_{\substack{A \\ |A|=r}} \epsilon(\underline{z}_0^{(\bar{A})}) . \quad (4)$$

The function u means here the average of the observations which constitute the argument of the function.

The following numerical values of the estimators are obtained if the data material above is used:

$n=4; \quad p=3$

1		1	1	0	0
0		1	0	1	0
0		1	0	0	0

$$(I) \quad |A| = |B| = 1$$

$$(i) \quad A = \{1\}$$

$$\underline{z}_0^{(\bar{A})} = (1)$$

$$\gamma(\underline{z}_0^{(\bar{A})}) = \frac{1+0}{2} = \frac{1}{2}$$

$$\epsilon(\underline{z}_0^{(\bar{A})}) = \frac{1+0+1+0+1}{5} = \frac{3}{5}$$

$$(ii) \quad A = \{2\}$$

$$\underline{z}_0^{(\bar{A})} = (0)$$

$$\gamma(\underline{z}_0^{(\bar{A})}) = \frac{\frac{1}{2}+0}{2} = \frac{1}{4}$$

$$\epsilon(\underline{z}_0^{(\bar{A})}) = \frac{\frac{1}{2}+0+\frac{1}{2}+0+\frac{1}{2}+\frac{1}{2}+0}{7} = \frac{2}{7}$$

$$(iii) \quad A = \{3\}$$

$$\underline{z}_0^{(\bar{A})} = (0)$$

$$\gamma(\underline{z}_0^{(\bar{A})}) = \frac{\frac{1}{2}+\frac{1}{2}+0}{3} = \frac{1}{3}$$

$$\epsilon(\underline{z}_0^{(\bar{A})}) = \frac{2}{7}$$

$$(iv) \quad r = 1$$

$$\gamma_1 = \frac{1}{\binom{p}{1}} \sum_{\substack{A \\ |A|=1}} \gamma(\underline{z}_0^{(\bar{A})})$$

$$\gamma_1 = \frac{1}{3} \left(\frac{1}{2} + \frac{1}{4} + \frac{1}{3} \right) = \frac{13}{36}$$

$$\epsilon_1 = \frac{1}{\binom{p}{1}} \sum_{\substack{A \\ |A|=1}} \epsilon(\underline{z}_0^{(\bar{A})})$$

$$\epsilon_1 = \frac{1}{3} \left(\frac{3}{5} + \frac{2}{7} + \frac{2}{7} \right) = \frac{41}{105}$$

$$(II) \quad |A| = |B| = 2$$

$$(i) \quad A = \{1, 2\}$$

$$\underline{z}_0^{(\bar{A})} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\gamma(\underline{z}_0^{(\bar{A})}) = \frac{0+0}{2} = 0$$

$$\epsilon(\underline{z}_0^{(\bar{A})}) = \frac{0+0+0+0}{4} = 0$$

$$(ii) \quad A = \{1, 3\}$$

$$\underline{z}_0^{(\bar{A})} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\gamma(\underline{z}_0^{(\bar{A})}) = \frac{0}{1} = 0$$

$$\epsilon(\underline{z}_0^{(\bar{A})}) = 0$$

$$(iii) \quad A = \{2, 3\}$$

$$\underline{z}_0^{(\bar{A})} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\gamma(\underline{z}_0^{(\bar{A})}) = \frac{1+0}{2} = \frac{1}{2}$$

$$\epsilon(\underline{z}_0^{(\bar{A})}) = \frac{0+1+0+1+0}{5} = \frac{2}{5}$$

$$(iv) \quad r = 2$$

$$\gamma_2 = \frac{1}{\binom{p}{2}} \sum_{\substack{A \\ |A|=2}} \gamma(\underline{z}_0^{(\bar{A})})$$

$$\gamma_2 = \frac{1}{3} (0 + 0 + \frac{1}{2}) = \frac{1}{6}$$

$$\epsilon_2 = \frac{1}{\binom{p}{2}} \sum_{\substack{A \\ |A|=2}} \epsilon(\underline{z}_0^{(\bar{A})})$$

$$\epsilon_2 = \frac{1}{3} (0 + 0 + \frac{2}{5}) = \frac{2}{15}$$

3.4 Estimation of the Bayes estimator with regression procedure

The EB estimators which are proposed in Chapter 3.2 are not smooth according to the definition in Maritz (1967), where an EB estimator is said to be smooth if it depends on all numerical values of the past observations. This is not fulfilled for the estimators of Chapter 3.2 because they do not use all numerical values of the past data explicitly. A smooth estimation method to estimate the Bayes estimator is the regression method described below. The regression method was used by Krutchkoff (1967) in the supplementary sample approach, cf. Chapter 2.2.

3.4.1 ϕ procedure

Suppose that the expected value of the posterior distribution for θ given x is a linear function of x . Thus

$$E(\theta|x) = \alpha + \beta x . \quad (1)$$

This assumption is usually not true, but it will be evident that this assumption can be used and lead to consequences of practical importance.

Consider the situation that n p -vectors of past and one p -vector of current data are present, i.e. the situation which is summed up in the tableau (3.1(5)). Let A be a subset of r index from the set $\Omega = \{1, \dots, p\}$, according to the definition in (3.2.1(34)). Suppose the function u is such that

$$E\{u(\underline{z}^{(A)})|\theta\} = \theta \quad (2)$$

where $\underline{z}^{(A)}$ as above, denotes the vector which remains when the r variables, whose indices are in A , have been omitted from the vector $\underline{z}$.

The following notation is introduced

$$x = v(\underline{z}^{(\bar{A})}) \quad (3)$$

$$y = u(\underline{z}^{(A)}) \quad (4)$$

where the function v has the same properties as u , defined above.

According to (3.2.1(6)) the Bayes estimator $\delta(x)$ can now be written as

$$\delta(x) = E(\theta|x) = E(y|x) \quad (5)$$

thus, the expression (1) can be expressed as

$$E(y|x) = \alpha_r + \beta_r x. \quad (6)$$

If we use the regression model

$$y = \alpha_r + \beta_r x + \epsilon \quad (7)$$

where ϵ is an error with zero mean and finite variance say σ^2 . The past values of x and y

$$x_i = v(\underline{z}_i^{(\bar{A})}) \quad (8)$$

$$i = 1, \dots, n$$

$$y_i = u(\underline{z}_i^{(A)}) \quad (9)$$

can then be considered as observations from a linear regression model with x as the independent variable. It is possible to obtain the usual least squares estimators of β_r and α_r as

$$b_A = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (10)$$

and

$$a_A = \bar{y} - b_A \bar{x} \quad (11)$$

where

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad (12)$$

and

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n} \quad (13)$$

Assuming that $(x_i - \bar{x})^2$ is not zero for all x_i , we can show that

$$\lim_{n \rightarrow \infty} \text{Var}\{a_A + b_A x | \underline{x}, x\} = 0 \quad (14)$$

where

$$\underline{x} = (x_1, \dots, x_n) \quad .$$

Let the non-parametric EB estimator $\phi(x_0)$ be defined as

$$\phi(x_0) = \begin{cases} a_A + b_A x_0 & \text{when } S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2 \neq 0 \\ 0 \text{ or } v(x_0) \text{ or } \bar{y} & \text{otherwise} \end{cases} \quad (15)$$

The extra risk of the estimator (15) becomes

$$\begin{aligned} r(\phi(x_0)) &= E_{Z, x_0} \{ \phi(x_0) - \delta(x_0) \}^2 = E_{x_0} E_{\underline{x} | x_0} E_{Z | \underline{x}, x_0} \{ \phi(x_0) - \delta(x_0) \}^2 = \\ &= E_{x_0} \{ E_{Z | \underline{x}, x_0} \{ (\phi(x_0) - \delta(x_0))^2 | \underline{x}, x_0, S_{xx} \neq 0 \} P(S_{xx} \neq 0) + \end{aligned} \quad (16)$$

(16) cont.

$$\begin{aligned}
 & + \sum_{\underline{z}} E_{\underline{x}_0} \{ (\phi(x_0) - \delta(x_0))^2 | \underline{z}, x_0, S_{xx} = 0 \} P(S_{xx} = 0) = \\
 & = E_{x_0} \{ \text{Var}\{a_A + b_A x_0 | \underline{z}, x_0, S_{xx} \neq 0\} P(S_{xx} \neq 0) \} + E_{x_0} \{ \delta^2(x_0) P(S_{xx} = 0) \}.
 \end{aligned}$$

It is true only for degenerate $f(x|\theta)$ that $S_{xx} = 0$ and we can thus see that

$$\lim_{n \rightarrow \infty} P(S_{xx} = 0) = 0 \quad (17)$$

while

$$\lim_{n \rightarrow \infty} E_{x_0} \{ \text{Var}\{a_A + b_A x_0 | \underline{z}, x_0, S_{xx} \neq 0\} P(S_{xx} \neq 0) \} \leq \quad (18)$$

$$\leq \lim_{n \rightarrow \infty} E_{x_0} \{ \lim_{n \rightarrow \infty} \text{Var}\{a_A + b_A x_0 | \underline{z}, x_0, S_{xx} \neq 0\} P(S_{xx} \neq 0) \} = 0.$$

The limit of the extra risk is

$$\lim_{n \rightarrow \infty} r(\phi(x_0)) = 0 \quad (19)$$

The estimator $\phi(x_0)$ is asymptotically optimal in the special case when the Bayes estimator is defined as $\delta(x_0)$, cf. the comments after the estimators (3.2.1(37), (38)). The estimator is also asymptotically optimal in the sense above if it is made equal to $\bar{y}$ or $v(\underline{z})$ instead of 0 when $S_{xx} = 0$.

For every r , $\binom{p}{r}$ different A can be constructed and it is possible to construct just as many estimators of the type (15). These estimators are combined by constructing the average

$$\phi_r = \frac{1}{\binom{p}{r}} \sum_{\substack{A \\ |A|=r}} \phi(x_0) = \sum_{\substack{A \\ |A|=r}} a_A / \binom{p}{r} + \sum_{\substack{A \\ |A|=r}} b_A x_0 / \binom{p}{r} \quad (20)$$

where

$$x_0 = v(\underline{z}_0^{(\bar{A})}) .$$

The estimator (20) is to be preferred to the estimator (15), because of the inequalities below.

$$E_Z W(\phi_r) \leq E_Z W(\phi(x)) \quad (21)$$

and

$$W(\delta_r) \leq W(\delta(x)) \quad (22)$$

where

$$\delta_r = \frac{1}{\binom{p}{r}} \sum_{\substack{A \\ |A|=r}} \delta(x); \quad x = v(\underline{z}^{(\bar{A})}) \quad (23)$$

and

$$\delta(x) = E(\theta|x) . \quad (24)$$

Similar inequalities have been used above to justify the average estimators of the γ and the ϵ procedures.

Now, as above in the Chapters 3.2.1 and 3.2.2, it is natural to ask if there is some optimal r or if an optimal combination of the $p - 1$ different estimators (20) can be constructed.

3.4.2 ψ and ω procedures

In the Chapter 3.2.2 we stated that the past data can be permuted in $\binom{p}{r}$ different ways for every constellation A of current data. This possibility will

also be used for the regression method. Let B have the same meaning as A cf. (3.2.1(34)). The past observations of x and y can be defined as

$$x_i = v(\underline{z}_i^{(\bar{B})}) \quad (1)$$

$$i = 1, \dots, n$$

$$y_i = u(\underline{z}_i^{(B)}) \quad (2)$$

The current observation of x is defined as

$$x_0 = v(\underline{z}_0^{(\bar{A})}) . \quad (3)$$

The estimator corresponding to the estimator (3.4.1(20)) becomes, when past data are permuted

$$\psi_r = \frac{1}{\binom{p}{r}} \sum_{\substack{A \\ |A|=r}} \frac{1}{\binom{p}{r}} \sum_{\substack{B \\ |B|=r}} \phi_B(x_0) \quad (4)$$

where B in $\phi_B(x_0)$ indicates which partition of the samples the estimates a and b are built up from. The expression (4) means that one set of $\binom{p}{r}$ estimates a and b are obtained from past data. These estimates a and b are the same for all the $\binom{p}{r}$ sets of the current observation, x_0 . The estimator

$$\frac{1}{\binom{p}{r}} \sum_{\substack{B \\ |B|=r}} \phi_B(x_0) \quad (5)$$

is asymptotically optimal, because all the estimators $\phi_B(x_0)$ which the sum contains are asymptotically optimal. This is true provided that the Bayes estimator is defined as $\delta(x_0)$. The estimator (4) is to be preferred to the estimator (5). By using the same thinking as before, we can justify this assertion, cf. the estimator (3.4.1(20)).

Since r can vary between 1 and $p-1$ the question arises, as above, if there

is some r which is optimal or if there is one combination of the estimators (4) which is optimal.

Another way of constructing past observations x and y , which will then be used to calculate the estimates a and b , is to let B take all the $\binom{p}{r}$ constellations in all the past samples, and from these values to construct the corresponding x and y values. Thus the past data become

$$x_{ij} = v(z_i^{(\bar{B})}) \quad i = 1, \dots, n \quad (6)$$

$$y_{ij} = u(z_i^{(B)}) \quad j = 1, \dots, \binom{p}{r} \quad (7)$$

The estimates a and b are then calculated from the $n \binom{p}{r}$ observations in (6) and (7). This means that one set of a and b is obtained for every r . The difference between this and the method used above in the estimator (4), is that here we get one set, and there we got $\binom{p}{r}$ sets of a and b for every r . The following estimator is constructed

$$\omega_r = \frac{1}{\binom{p}{r}} \sum_{\substack{A \\ |A|=r}} \phi(x_0) = a_r + b_r \sum_{\substack{A \\ |A|=r}} x_0 / \binom{p}{r} \quad (8)$$

where

$$x_0 = v(z_0^{(\bar{A})}) .$$

The observations built up in (6) and (7) are certainly not independent but it will be seen from Monte Carlo simulations that the estimator (8) has desirable properties.

3.4.3 Choice of one estimator among $p - 1$ estimators of the same type

All the estimators γ_r and $\gamma_r^{(c)}$ (3.2.1(39), (40)), ϵ_r and $\epsilon_r^{(c)}$ (3.2.2(16), (17)), c_r (3.4.1(20)) and ϵ_r and ω_r (3.4.2(4), (8)) depend among other things

on r , where r is the number of elements in A defined in (3.2.1(34)). Since r can take any integral value between 1 and $p - 1$ it is possible to construct $p - 1$ estimators for each of the procedures mentioned above. A question which arises is if there is some optimal value of r , and thus if the corresponding estimator has the least expected risk of the $p - 1$ estimators. In the search for an optimal r , it is possible to make a theoretical analysis of the regression estimators, when the conditional distribution, $f(z|\theta)$, and the prior distribution, $G(\theta)$, are known.

Let

$$\rho(x) = \alpha_r + \beta_r x \quad (1)$$

be the regression estimator of $E(\theta|v(\underline{z}))$ when $|A| = r$ and when x is defined as in (3.4.1(3)). It is possible to construct $\binom{p}{r}$ different A and thus the same number of estimators (1). It is natural to construct the average of all these $\binom{p}{r}$ estimators as

$$\rho_r = \frac{1}{\binom{p}{r}} \sum_{\substack{A \\ |A|=r}} \rho(x) = \alpha_r + \beta_r \frac{1}{\binom{p}{r}} \sum_{\substack{A \\ |A|=r}} v(\underline{z}^{(\bar{A})}) . \quad (2)$$

The parameters α_r and β_r can be expressed as

$$\begin{cases} \beta_r = \frac{\text{Cov}(x, y)}{\text{Var}(x)} = \frac{\text{Cov}(E(x|\theta), E(y|\theta)) + E \text{Cov}(x, y|\theta)}{\text{Var } E(x|\theta) + E \text{Var}(x|\theta)} \\ \alpha_r = (1 - \beta_r) E x = (1 - \beta_r) E E(x|\theta) \end{cases} \quad (3)$$

where x and y are defined in (3.4.1(3) and (4) respectively). Let the conditional distribution, $f(z|\theta)$, be Bernoulli(θ) and the prior distribution, $G(\theta)$, be Beta(e, f). Let μ and σ^2 be the expected value and the variance respectively of the prior distribution, $G(\theta)$. The following equalities can now be stated

$$\frac{1}{\binom{p}{r}} \sum_{\substack{A \\ |A|=r}} v(\underline{z}^{(\bar{A})}) = \bar{z} = \frac{\sum_{i=1}^p z_i}{p} \quad (4)$$

$$E(x|\theta) = \theta \quad (5)$$

$$E(y|\theta) = \theta \quad (6)$$

$$\text{Var}(x|\theta) = \frac{\theta(1-\theta)}{r} \quad (7)$$

$$\text{Cov}(x, y|\theta) = 0 \quad (8)$$

The expressions above are valid if the functions u and v denote the ML estimator. The parameters β_r and α_r can be expressed in terms of μ and σ^2 as

$$\beta_r = \frac{\sigma^2}{\sigma^2 + E \frac{\theta(1-\theta)}{r}} = \frac{\sigma^2}{\sigma^2 + \frac{1}{r} (\mu - \sigma^2 - \mu^2)} \quad (9)$$

$$\alpha_r = (1 - \beta_r)\mu \quad (10)$$

If

$$\tau = \frac{\mu - \sigma^2 - \mu^2}{\sigma^2} \quad (11)$$

the estimator (2) can be expressed as

$$\rho_r = \mu - \mu \frac{1}{1 + \frac{\tau}{r}} + \bar{z} \frac{1}{1 + \frac{\tau}{r}} \quad (12)$$

If τ in (11) is expressed in terms of the parameters e and f of the prior distribution we get

$$\tau = e + f. \quad (13)$$

Finally the estimator (12) can be expressed as

$$\rho_r = \frac{e}{e+f} - \frac{e}{e+f} \frac{1}{1 + \frac{e+f}{r}} + \bar{z} \frac{1}{1 + \frac{e+f}{r}} = \frac{e + r\bar{z}}{r + e + f}. \quad (14)$$

The Bayes estimator $\delta(\bar{z})$ is, when $f(z|\theta)$ is Bernoulli(θ) and $G(\theta)$ is Beta(e, f), given by (cf. Maritz (1970))

$$\delta(\bar{z}) = \frac{e + p\bar{z}}{p + e + f}. \quad (15)$$

If the expressions (14) and (15) are compared with each other we can see that ρ_p equals the Bayes estimator. Thus the extra risk

$$E\{\rho_r - \delta(\bar{z})\}^2 \quad (16)$$

is minimized when $r = p$. Now, r can just take the integers $1, 2, \dots, p-1$ and since (16) is a monotone function for these values, (16) is minimized for $r = p - 1$. This means that x shall be built up out of $p - 1$ observations and y of 1 observation, if the extra risk (16) is to be as small as possible.

Before we formulate a general rule of choosing r in the regression model we will give another example.

Now, let the conditional distribution, $f(z|\theta)$, be $N(\theta, 1)$ and the prior distribution, $G(\theta)$, be $N(\mu, \sigma^2)$. As above, we can establish the following

$$\frac{1}{\binom{p}{r}} \sum_{\substack{A \\ |A|=r}} v(\underline{z}^{(\bar{A})}) = \bar{z} = \frac{\sum_{i=1}^p z_i}{p} \quad (17)$$

$$E(x|\theta) = \theta \quad (18)$$

$$E(y|\theta) = \theta \quad (19)$$

$$\text{Var}(x|\theta) = \frac{1}{r} \quad (20)$$

$$\text{Cov}(x, y|\theta) = 0 \quad (21)$$

As before, the expressions above are valid if the functions u and v denote the ML estimator. The parameter β_r and α_r can now be expressed as

$$\beta_r = \frac{\sigma^2}{\sigma^2 + \frac{1}{r}} \quad (22)$$

$$\alpha_r = (1 - \beta_r)\mu \quad (23)$$

By using (22) and (23), the estimator (2) can be written as

$$\rho_r = \mu - \frac{\sigma^2 r}{\sigma^2 r + 1} \mu + \frac{\sigma^2 r}{\sigma^2 r + 1} \bar{z} = \frac{\mu + \sigma^2 r \bar{z}}{\sigma^2 r + 1} \quad (24)$$

The Bayes estimator $\delta(\bar{z})$ can be constructed when the conditional distribution, $f(z|\theta)$, is $N(\theta, 1)$ and the prior distribution, $G(\theta)$, is $N(\mu, \sigma^2)$ and becomes

$$\delta(\bar{z}) = \frac{\mu + \sigma^2 p \bar{z}}{\sigma^2 p + 1} \quad (25)$$

e.g. cf. Maritz (1970). We see that ρ_p also in this case equals the Bayes estimator. Thus, we must choose r equal to $p - 1$ to get the extra risk (16) as small as possible.

Is it generally true that the optimal value of r equals $p - 1$? No, this is unfortunately not the case. It is true, however, if the regression of θ on x is linear. This means that the model of the regression procedure is true which implies that $E(\theta | v(\underline{z})) = \rho_p$. If in addition it is true that the extra risk (16) is a monotone function of r for the integers $1, 2, \dots, p-1$, the optimal value of r equals $p - 1$.

The regression of θ on x is linear for some of the naturally conjugate distributions e.g. Bernoulli-Beta, Poisson-Gamma, Normal-Normal with the expected value unknown and Normal- χ^2 with the variance unknown. The Bayes estimator of the three first cases can be found e.g. in Maritz (1970) and for the last case in e.g. Lindley (1965). In the three first cases the functions u and v denote the average of $p-r$ and r observations z , respectively, but in the last case the average of $p-r$ and r observations $(z - \mu)^2$, respectively, where μ is a known mean value.

In practice one has at best only a slight idea of the distributions $f(z | \theta)$ and $G(\theta)$. Thus one does not know anything about the regression of θ on x and because of that no optimal r can be determined in advance with the method presented in this Chapter. Moreover, the parameters α_r and β_r require to be estimated. The r -value which is then optimal in the EB estimator $\rho_r(\underline{z}_0, Z)$ is not investigated here. It is not necessarily true that the best r -value is $p - 1$ in this case, since the estimators of α_r and β_r also are affected by r .

4. THE QUALITY CONTROL PROBLEM

In a manufacturing process a large number of units of a certain product is produced every day. It can be bolts, electric bulbs, transistors and so on. In for example every consignment or every day's production or any other batch of items, we want to estimate the quality of the product. By the quality is here meant the proportion of defects in the batch. The estimation of the quality is carried out by randomly drawing out a small sample from a consignment, day's production or from any other batch of items in which the quality is to be estimated. Every unit of the sample is checked to see whether it is defective or not. From the sample proportion of defects, it is possible to estimate the batch proportion of defects, θ . If we let the day's production be the batch of interest, the θ will vary from day to day. That the batch proportion of defects, θ , is varying from day to day can depend on e.g. the machine handling, the air humidity, the temperature, the raw material and so on. Because the parameter θ is expected to vary from day to day only the sample of today can be used to estimate the parameter, θ_0 , of today with traditional methods. This is not the case with EB procedures, where also the past samples can be used to estimate the parameter, θ_0 , of today. If EB procedures are used, we expect, under certain conditions, to get more confident estimators, or equivalently less samples can be drawn with maintained accuracy in the estimators compared with the use of traditional estimators.

If EB procedures are used to estimate θ_0 , in the type of quality control described above, the conditional distribution, $f(z|\theta)$, will be known. We are thus not forced to use non-parametric EB procedures, but it will be clear in Chapter 5 that the non-parametric EB procedures vindicate themselves well compared to the EB procedures which use the knowledge of the conditional distribution, $f(z|\theta)$. For this comparison on the whole to be carried out the conditional distribution, $f(z|\theta)$, of course has to be known.

4.1 Notation

The notations defined in Chapter 3 will be used also in Chapter 4. Certain new assumptions will be made. The conditional distribution $f(z|\theta)$ will be Bernoulli(θ). A defect unit is denoted by 1 and a correct unit by 0. For every

value of the parameter, θ , the p observations in the samples will be p independent observations of a Bernoulli(θ) variable. The general estimation functions u and v in Chapter 3 will here denote the ML estimator of θ , thus the average of the observations which the argument of u or v consists of.

4.2 Five non-parametric EB procedures

The five non-parametric EB procedures which have been treated in Chapter 3 will in this chapter be presented when the conditional distribution $f(z|\theta)$ is Bernoulli(θ).

4.2.1 γ procedure

For the non-parametric EB procedure γ , which is described in Chapter 3.2.1, certain simplifications can be done in the Bernoulli case. If we consider the estimator (3.2.1(37))

$$\gamma(\underline{z}_0^{(\bar{A})}) = \begin{cases} \frac{\sum_{i=1}^n u(\underline{z}_i^{(A)}) J(\underline{z}_i^{(\bar{A})} = \underline{z}_0^{(\bar{A})})}{m(\underline{z}^{(\bar{A})}, \underline{z}_0^{(\bar{A})})} & \text{if } m(\underline{z}^{(\bar{A})}, \underline{z}_0^{(\bar{A})}) > 0 \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

then $u(\underline{z}_i^{(A)})$ is the average of the p - r observations, with index which is in $\bar{A}$, of the i th past sample. The indicator

$$J(\underline{z}_i^{(\bar{A})} = \underline{z}_0^{(\bar{A})}) \quad (2)$$

is defined in (3.2.1(28)). There we can see, that the r observations of the vector, $\underline{z}_i^{(\bar{A})}$, in the general case have to be permuted to determine if $J(\cdot)$ is to be 0 or 1. In the Bernoulli case when $z = 0, 1$ it is sufficient to sum up the observations of the vector, $\underline{z}_i^{(\bar{A})}$, and then determine if this sum equals the sum of the observations of the vector $\underline{z}_0^{(\bar{A})}$. Since no permutation

has to be done we can write the indicator (2) as

$$I(v(\underline{z}_i^{(\bar{A})}) = v(\underline{z}_0^{(\bar{A})})) . \quad (3)$$

where the indicator $I(\cdot)$ is defined in page 43.

4.2.2 ϵ procedure

The non-parametric EB procedure, ϵ , described in Chapter 3.2.2 can be simplified when the conditional distribution $f(z|\theta)$ is Bernoulli(θ). In order to calculate the estimates ϵ we only need to know the number of ones, thus the sum, of each sample and not as in the γ procedure the sample. The notation s_i denotes the sum of the observations of the vector $\underline{z}_i$ and s denotes the sum of the observations of the vector $\underline{z}_0$.

Let us consider the situation when r equals one, thus the vector $\underline{z}_0^{(\bar{A})}$ contains just one observation. Two cases can be distinguished, either the vector $\underline{z}_0^{(\bar{A})}$ is (0) or (1). If the notations of (3.2.2(16)) are used we obtain

$$\epsilon(\underline{z}_0^{(\bar{A})}) = \begin{cases} \frac{\sum_B \sum_{i=1}^n u(\underline{z}_i^{(B)}) J(\underline{z}_i^{(\bar{B})} = \underline{z}_0^{(\bar{A})})}{\sum_B \sum_{i=1}^n J(\underline{z}_i^{(\bar{B})} = \underline{z}_0^{(\bar{A})})} & \text{if } |B|=r \\ \sum_B \sum_{i=1}^n J(\underline{z}_i^{(\bar{B})} = \underline{z}_0^{(\bar{A})}) > 0 \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

The first case

$$\underline{z}_0^{(\bar{A})} = (0) . \quad (i)$$

If the denominator

$$\sum_B \sum_{i=1}^n J(\underline{z}_i^{(\bar{B})} = \underline{z}_0^{(\bar{A})})$$

$|B|=r$

of the expression (1) does not equal zero we obtain

$$\epsilon(\underline{z}_0^{(\bar{A})}) = \frac{\sum_{j=1}^p \sum_{i=1}^n s_i I(z_{ij} = 0)}{(p-1) \sum_{j=1}^p \sum_{i=1}^n I(z_{ij} = 0)} = \frac{\sum_{i=1}^n s_i \sum_{j=1}^p (1 - z_{ij})}{(p-1) \sum_{j=1}^p \sum_{i=1}^n (1 - z_{ij})} =$$

$$= \frac{\sum_{i=1}^n s_i (p - s_i)}{n \sum_{i=1}^n (p - s_i)} \quad (2)$$

where $\frac{s_i}{p-1}$ is $u(\underline{z}_i^{(B)})$ of the expression (1).

The second case

$$\underline{z}_0^{(\bar{A})} = (1) \quad (ii)$$

If the denominator

$$\sum_{B \atop |B|=r} \sum_{i=1}^n J(\underline{z}_i^{(\bar{B})}) = \underline{z}_0^{(\bar{A})}$$

of the expression (1) does not equal zero we obtain

$$\epsilon(\underline{z}_0^{(\bar{A})}) = \frac{\sum_{j=1}^p \sum_{i=1}^n (s_i - 1) I(z_{ij} = 1)}{(p-1) \sum_{j=1}^p \sum_{i=1}^n I(z_{ij} = 1)} = \frac{\sum_{i=1}^n (s_i - 1) \sum_{j=1}^p z_{ij}}{(p-1) \sum_{j=1}^p \sum_{i=1}^n z_{ij}} =$$

$$= \frac{\sum_{i=1}^n s_i (s_i - 1)}{n \sum_{i=1}^n s_i} \quad (3)$$

When r equals one it is possible to construct p different A . In the Bernoulli case corresponding to the average estimator (3.2.2(16)) we have

$$\epsilon_1 = \begin{cases} \frac{1}{p} \left\{ (p-s) \frac{\sum_{i=1}^n s_i (p-s_i)}{(p-1) \sum_{i=1}^n (p-s_i)} + s \frac{\sum_{i=1}^n s_i (s_i-1)}{(p-1) \sum_{i=1}^n s_i} \right\}, & \text{if} \\ & (p-s) \sum_{i=1}^n (p-s_i) + s \sum_{i=1}^n s_i > 0 \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

and corresponding to the modified average estimator (3.2.2(17)) we have

$$\epsilon_1^{(c)} = \begin{cases} \epsilon_1 & \text{if } (p-s) \sum_{i=1}^n (p-s_i) + s \sum_{i=1}^n s_i > c \\ \frac{s}{p} & \text{if } (p-s) \sum_{i=1}^n (p-s_i) + s \sum_{i=1}^n s_i \leq c \end{cases} \quad (5)$$

In the expression (4), the value $0/0$ can arise. This value will, here and subsequently be defined as zero.

If r equals 2, three cases can be distinguished.

The first case

$$\underline{z}_0^{(\bar{A})} = (0 \ 0)' . \quad (i)$$

If the denominator

$$\sum_B \sum_{i=1}^n J(\underline{z}_i^{(\bar{B})}) = \underline{z}_0^{(\bar{A})} \\ |B|=r$$

of the expression (1) does not equal zero we obtain

$$\begin{aligned}
 \epsilon \left(\frac{\bar{A}}{z_0} \right) &= \frac{\sum_{j < j'}^p \sum_{i=1}^p \sum_{i=1}^n s_i I(z_{ij} + z_{ij'} = 0)}{(p-2) \sum_{j < j'}^p \sum_{i=1}^p \sum_{i=1}^n I(z_{ij} + z_{ij'} = 0)} = \\
 &= \frac{\sum_{j \neq j'}^p \sum_{i=1}^p \sum_{i=1}^n s_i I(z_{ij} + z_{ij'} = 0)}{(p-2) \sum_{j \neq j'}^p \sum_{i=1}^p \sum_{i=1}^n I(z_{ij} + z_{ij'} = 0)} = \\
 &= \frac{\sum_{i=1}^n s_i \sum_{j \neq j'}^p \sum_{j'=1}^p (1 - z_{ij})(1 - z_{ij'})}{(p-2) \sum_{i=1}^n \sum_{j \neq j'}^p \sum_{j'=1}^p (1 - z_{ij})(1 - z_{ij'})} = \\
 &= \frac{\sum_{i=1}^n s_i \sum_{j=1}^p \sum_{j'=1}^p (1 - z_{ij})(1 - z_{ij'}) - \sum_{i=1}^n s_i \sum_{j=1}^p (1 - z_{ij})}{(p-2) \left\{ \sum_{i=1}^n \sum_{j=1}^p \sum_{j'=1}^p (1 - z_{ij})(1 - z_{ij'}) - \sum_{i=1}^n \sum_{j=1}^p (1 - z_{ij}) \right\}} = \\
 &= \frac{\sum_{i=1}^n s_i (p - s_i)(p - s_i - 1)}{(p-2) \sum_{i=1}^n (p - s_i)(p - s_i - 1)} .
 \end{aligned} \tag{6}$$

The second case

$$\frac{\bar{A}}{z_0} = (0 \ 1)' \text{ or } (1 \ 0)' . \tag{ii}$$

If the denominator

$$\sum_{\substack{B \\ |B|=r}} \sum_{i=1}^n J(\underline{z}_i^{(\bar{B})} = \underline{z}_0^{(\bar{A})})$$

of the expression (1) does not equal zero we obtain

$$\epsilon(\underline{z}_0^{(\bar{A})}) = \frac{\sum_{j < j'}^p \sum_{i=1}^p \sum_{i=1}^n (s_i - 1) I(z_{ij} + z_{ij'} = 1)}{(p-2) \sum_{j < j'}^p \sum_{i=1}^p \sum_{i=1}^n I(z_{ij} + z_{ij'} = 1)} = \quad (7)$$

$$= \frac{\sum_{j \neq j'}^p \sum_{i=1}^p \sum_{i=1}^n (s_i - 1) I(z_{ij} + z_{ij'} = 1)}{(p-2) \sum_{j \neq j'}^p \sum_{i=1}^p \sum_{i=1}^n I(z_{ij} + z_{ij'} = 1)} =$$

$$= \frac{\sum_{i=1}^n (s_i - 1) \sum_{j \neq j'}^p \sum_{j'=1}^p ((1 - z_{ij})z_{ij'} + (1 - z_{ij'})z_{ij})}{(p-2) \sum_{i=1}^n \sum_{j \neq j'}^p \sum_{j'=1}^p ((1 - z_{ij})z_{ij'} + (1 - z_{ij'})z_{ij})} =$$

$$= \left\{ \sum_{i=1}^n (s_i - 1) \sum_{j=1}^p \sum_{j'=1}^p ((1 - z_{ij})z_{ij'} + (1 - z_{ij'})z_{ij}) \right\} -$$

$$- \sum_{i=1}^n (s_i - 1) \sum_{j=1}^p (1 - z_{ij})z_{ij} \} / \{(p-2) \cdot$$

$$\cdot \left(\sum_{i=1}^n \sum_{j=1}^p \sum_{j'=1}^p ((1 - z_{ij})z_{ij'} + (1 - z_{ij'})z_{ij}) - \right.$$

(7) cont.

$$\begin{aligned}
 & - \sum_{i=1}^n \sum_{j=1}^p (1 - z_{ij}) z_{ij} \} = \\
 & = \frac{\sum_{i=1}^n s_i (s_i - 1) (p - s_i)}{(p-2) \sum_{i=1}^n s_i (p - s_i)} .
 \end{aligned}$$

The third case

$$\underline{z}_0^{(\bar{A})} = (1 \ 1)' . \quad (\text{iii})$$

If the denominator

$$\begin{aligned}
 & \sum_{\substack{B \\ |B|=r}} \sum_{i=1}^n J(\underline{z}_i^{(\bar{B})} = \underline{z}_0^{(\bar{A})}) \\
 & \quad |B|=r
 \end{aligned}$$

of the expression (1) does not equal zero we obtain

$$\begin{aligned}
 e(\underline{z}_0^{(\bar{A})}) &= \frac{\sum_{j < j'}^p \sum_{i=1}^p \sum_{i=1}^n (s_i - 2) I(z_{ij} + z_{ij'} = 2)}{(p-2) \sum_{j < j'}^p \sum_{i=1}^p \sum_{i=1}^n I(z_{ij} + z_{ij'} = 2)} = \quad (8)
 \end{aligned}$$

$$\begin{aligned}
 & \sum_{j \neq j'}^p \sum_{i=1}^p \sum_{i=1}^n (s_i - 2) I(z_{ij} + z_{ij'} = 2) \\
 &= \frac{\sum_{j \neq j'}^p \sum_{i=1}^p \sum_{i=1}^n I(z_{ij} + z_{ij'} = 2)}{(p-2) \sum_{j \neq j'}^p \sum_{i=1}^p \sum_{i=1}^n I(z_{ij} + z_{ij'} = 2)} =
 \end{aligned}$$

cont. next page

(8) cont.

$$\begin{aligned}
&= \frac{\sum_{i=1}^n (s_i - 2) \sum_{j \neq j'}^p z_{ij} z_{ij'}}{(p-2) \sum_{i=1}^n \sum_{j \neq j'}^p z_{ij} z_{ij'}} = \\
&= \frac{\sum_{i=1}^n (s_i - 2) \sum_{j=1}^p \sum_{j'=1}^p z_{ij} z_{ij'} - \sum_{i=1}^n (s_i - 2) \sum_{j=1}^p z_{ij}^2}{(p-2) \left(\sum_{i=1}^n \sum_{j=1}^p \sum_{j'=1}^p z_{ij} z_{ij'} - \sum_{i=1}^n \sum_{j=1}^p z_{ij}^2 \right)} = \\
&= \frac{\sum_{i=1}^n s_i (s_i - 1) (s_i - 2)}{(p-2) \sum_{i=1}^n s_i (s_i - 1)}.
\end{aligned}$$

When r equals 2 it is possible to construct $\binom{p}{2}$ different A and the correspondence to the average estimator (3.2.2(16)) is in the Bernoulli case equal to

$$e_2 = \begin{cases} e_2 & \text{if } m_2 > 0 \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

where

$$e_2 = \frac{1}{\binom{p}{2}} \left\{ \frac{(p-s)(p-s-1)}{2} \frac{\sum_{i=1}^n s_i (p-s_i)(p-s_i-1)}{(p-2) \sum_{i=1}^n (p-s_i)(p-s_i-1)} + \right. \quad (10)$$

(10) cont.

$$+ s(p-s) \frac{\sum_{i=1}^n s_i(s_i-1)(p-s_i)}{(p-2) \sum_{i=1}^n s_i(p-s_i)} + \frac{s(s-1)}{2} \frac{\sum_{i=1}^n s_i(s_i-1)(s_i-2)}{(p-2) \sum_{i=1}^n s_i(s_i-1)} \}$$

and where

$$m_2 = \frac{(p-s)(p-s-1) \sum_{i=1}^n (p-s_i)(p-s_i-1)}{2 \cdot 2} + \quad (11)$$

$$+ s(p-s) \sum_{i=1}^n s_i(p-s_i) + \frac{s(s-1) \sum_{i=1}^n s_i(s_i-1)}{2 \cdot 2}.$$

It is possible to continue the deduction of the estimators, e_r , when $r = 3, 4, \dots, p-1$ in the way indicated above. A simpler way of deducing the estimator would be preferred. It is also possible to deduce a general expression of the estimator, e_r , for an arbitrary r in a much simpler way with aid of a combinatorial argument.

Let us consider the expression

$$\sum_{j < j'}^p \sum_{i=1}^p (1 - z_{ij})(1 - z_{ij'})$$

in (6). This expression is equal to the number of ways it is possible to draw, without replacement, two zeroes from the i th sample. There are $p - s_i$ zeroes of the i th sample and the first zero drawn can thus be chosen in $p - s_i$ different ways while the second zero can be chosen in $p - s_i - 1$ different ways. Thus, two zeroes can be chosen from the i th sample in $(p - s_i)(p - s_i - 1)$ different ways. In this expression we take account of the order of the elements. Since the different A are constructed without regarding the order of the index of A ,

the expression

$$\sum_{j < j'}^p \sum_{ij}^p (1 - z_{ij})(1 - z_{ij'})$$

is equal to

$$(p - s_i)(p - s_i - 1)/2 .$$

If we consider another expression e.g.

$$\sum_{j < j'}^p \sum_{ij}^p ((1 - z_{ij})z_{ij'} + (1 - z_{ij'})z_{ij})$$

which is part of (7), we can determine that this expression is equal to the number of ways it is possible to choose a zero and a one from the i th sample. This sample contains $p - s_i$ zeroes and s_i ones. The expression above is thus equal to

$$s_i(p - s_i) .$$

This argument can easily be extended to an arbitrary r . The average estimators which in the general case is given in (3.2.2(16)) can in the Bernoulli case be expressed as

$$\epsilon_r = \begin{cases} (4) & \text{if } r = 1 \\ e_r & \text{if } m_r > 0 \text{ and } r > 1 \\ 0 & \text{otherwise} \end{cases} \quad (12)$$

where

$$\begin{aligned}
e_r = & \frac{1}{(p)_r} \left\{ (p-s)_r \frac{\sum_{i=1}^n (p-s_i)_r s_i}{(p-r) \sum_{i=1}^n (p-s_i)_r} + \right. \\
& + \sum_{j=1}^{r-1} \binom{r}{j} (p-s)_{r-j} (s)_j \frac{\sum_{i=1}^n (p-s_i)_{r-j} (s_i)_{j+1}}{(p-r) \sum_{i=1}^n (p-s_i)_{r-j} (s_i)_j} + \\
& \left. + (s)_r \frac{\sum_{i=1}^n (s_i)_{r+1}}{(p-r) \sum_{i=1}^n (s_i)_r} \right\}
\end{aligned} \tag{13}$$

and where

$$m_r = \frac{1}{(r!)^2} \left\{ (p-s)_r \sum_{i=1}^n (p-s_i)_r + \right. \tag{14}$$

$$\left. + \sum_{j=1}^{r-1} \binom{r}{j} (p-s)_{r-j} (s)_j \sum_{i=1}^n \binom{r}{j} (p-s_i)_{r-j} (s_i)_j + (s)_r \sum_{i=1}^n (s_i)_r \right\}.$$

In (13) and (14) the expression $(n)_k$ is used and it is defined as

$$(n)_k = \frac{n!}{(n-k)!} = n(n-1) \cdots (n-k+1) . \tag{15}$$

Since $0/0$ is defined as 0 there is no need for the expression '0 otherwise' in (12). The reason why it nevertheless is there, is that we perhaps do not want to set the estimator (12) equal to zero when $m_r = 0$, but instead wish to set the estimator equal to e.g. s/p in this particular case. The same indication is valid also when r equals one in (4).

A general expression for the modified average estimator (3.2.2(17)), in the Bernoulli case, is

$$\epsilon_r^{(c)} = \begin{cases} (5) & \text{if } r = 1 \\ e_r & \text{if } m_r > c \text{ and } r > 1 \\ s/p & \text{if } m_r \leq c \text{ and } r > 1 \end{cases} \quad (16)$$

where e_r is defined in (13) and m_r in (14).

4.2.3 ϕ procedure

No particular simplifications can be made in the Bernoulli case to the regression estimator, ϕ , described in Chapter 3.4.1. It can, however, be pointed out that since the functions u and v here are the ML estimator, the expression (3.4.1(8))

$$x_i = v(\underline{z}_i^{(\bar{A})}) \quad i = 1, \dots, n \quad (1)$$

is equal to the average of the observations of the i th sample with index which is in A . While the expression (3.4.1(9))

$$y_i = u(\underline{z}_i^{(A)}) \quad i = 1, \dots, n \quad (2)$$

is equal to the average of the remaining observations of the i th sample.

4.2.4 Identities between the γ and ϕ procedures

Under certain conditions, in the Bernoulli case, the estimator, defined in (3.2.1(39)) is equal to the estimator ϕ , defined in (3.4.1(20)). In order to prove this, it is convenient to introduce the notation

$$t_j = \sum_{i=1}^n z_{ij} \quad (1)$$

which denotes the sum of the j th elements of the past samples, $\underline{z}_1, \dots, \underline{z}_n$.

The estimator γ_r defined in (3.2.1(39)) is built up as the sum of $\binom{p}{r}$ estimators $\gamma(\underline{z}_0^{(\bar{A})})$ defined in (3.2.1(37)). When r equals one, $\bar{A}$ contains just one index, say j and the last mentioned estimator can, if $m(\underline{z}^{(\bar{A})}, \underline{z}_0^{(\bar{A})}) > 0$ and if $t_j \neq 0$, be written as

$$\gamma(\underline{z}_0^{(\bar{A})}) = \frac{\sum_{i=1}^n s_i(1 - z_{ij})}{(p-1)(n - t_j)} (1 - z_{0j}) + \frac{\sum_{i=1}^n (s_i - 1)z_{ij}}{(p-1)t_j} z_{0j} = \quad (2)$$

$$= \frac{\sum_{i=1}^n s_i - \sum_{i=1}^n z_{ij} s_i}{(p-1)(n - t_j)} +$$

$$+ \frac{(n - t_j)(\sum_{i=1}^n z_{ij} s_i - \sum_{i=1}^n z_{ij}) - t_j(\sum_{i=1}^n s_i - \sum_{i=1}^n z_{ij} s_i)}{(p-1)(n - t_j)t_j} z_{0j} =$$

$$= \frac{\sum_{i=1}^n s_i - \sum_{i=1}^n z_{ij} s_i}{(p-1)(n - t_j)} + \{n \sum_{i=1}^n z_{ij} s_i - n \sum_{i=1}^n z_{ij} - t_j \sum_{i=1}^n z_{ij} s_i + t_j \sum_{i=1}^n z_{ij} -$$

$$- t_j \sum_{i=1}^n s_i + t_j \sum_{i=1}^n z_{ij} s_i\} / \{(p-1)(n - t_j)t_j\} z_{0j} = \frac{\sum_{i=1}^n s_i - \sum_{i=1}^n z_{ij} s_i}{(p-1)(n - t_j)} +$$

$$+ \frac{n(\sum_{i=1}^n z_{ij} s_i - \sum_{i=1}^n z_{ij}) - t_j(\sum_{i=1}^n s_i - \sum_{i=1}^n z_{ij})}{(p-1)(n - t_j)t_j} z_{0j}.$$

The average estimator defined in (3.2.1(39)) can, when r equals one, be written as

$$\gamma_1 = \frac{1}{p} \sum_{\substack{A \\ |A|=1}} \gamma(z_0^{(\bar{A})}) \quad (3)$$

Just as for the γ procedure, the estimator ϕ_r defined in (3.4.1(20)) is built up as the sum of $\binom{p}{r}$ estimators $\phi(x_0)$ defined in (3.4.1(15)). If r equals one, A contains index j and if

$$S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2 \neq 0$$

which is equivalent to $t_j \neq 0, n$, the last mentioned estimator can be expressed as

$$\phi(x_0) = a_A + b_A x_0 = a_A + b_A z_{0j} \quad (4)$$

where

$$b_A = \frac{\sum_{i=1}^n x_i y_i - \frac{1}{n} \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sum_{i=1}^n x_i^2 - \left(\frac{1}{n} \sum_{i=1}^n x_i \right)^2} = \frac{\sum_{i=1}^n z_{ij} (s_i - z_{ij}) - \frac{1}{n} \sum_{i=1}^n z_{ij} \sum_{i=1}^n (s_i - z_{ij})}{(p-1) \left(\sum_{i=1}^n z_{ij}^2 - \left(\frac{1}{n} \sum_{i=1}^n z_{ij} \right)^2 \right)} =$$

$$= \frac{\sum_{i=1}^n z_{ij} s_i - \sum_{i=1}^n z_{ij}^2 - t_j \left(\sum_{i=1}^n s_i - \sum_{i=1}^n z_{ij} \right)}{(p-1)(n - t_j)t_j} \quad (5)$$

and

$$a_A = \bar{y} - b_A \bar{x} = \frac{\sum_{i=1}^n (s_i - z_{ij})}{n(p-1)} - b_A \frac{\sum_{i=1}^n z_{ij}}{n} = \quad (6)$$

$$\begin{aligned} &= \frac{\sum_{i=1}^n (s_i - t_j)}{n(p-1)} - \frac{n(\sum_{i=1}^n z_{ij} s_i - \sum_{i=1}^n z_{ij}) - t_j(\sum_{i=1}^n s_i - \sum_{i=1}^n z_{ij})}{(p-1)(n - t_j)t_j} \frac{t_j}{n} = \\ &= \frac{1}{n(p-1)} \left\{ n \sum_{i=1}^n s_i - n t_j - t_j \sum_{i=1}^n s_i + t_j^2 - n \sum_{i=1}^n z_{ij} s_i + n \sum_{i=1}^n z_{ij} + \right. \\ &\quad \left. + t_j \sum_{i=1}^n s_i - t_j \sum_{i=1}^n z_{ij} \right\} / \{n - t_j\} = \frac{\sum_{i=1}^n s_i - \sum_{i=1}^n z_{ij} s_i}{(p-1)(n - t_j)} . \end{aligned}$$

If we now put the expressions (5) and (6) in the estimator (4) we obtain

$$\phi(x_0) = \frac{\sum_{i=1}^n s_i - \sum_{i=1}^n z_{ij} s_i}{(p-1)(n - t_j)} + \quad (7)$$

$$+ \frac{n(\sum_{i=1}^n z_{ij} s_i - \sum_{i=1}^n z_{ij}) - t_j(\sum_{i=1}^n s_i - \sum_{i=1}^n z_{ij})}{(p-1)(n - t_j)t_j} z_{0j} .$$

The average estimator defined in (3.4.1(20)) can, if r equals one, be expressed as

$$\phi_1 = \frac{1}{p} \sum_A \phi(x_0) \quad (8)$$

$|A| = 1$

If the expressions (2) and (7) are compared, we see that the expressions are identical. The estimators (3) and (8) are with two exceptions identical. The exceptions are when t_j equals 0 or n , as below.

If $\frac{z}{z_0}(\bar{A})$ is calculated from the first member of the expression (2) and $0/0$ is defined as 0 the estimators (2) and (7) will be equal also when $t_j = 0$ and $z_{0j} = 0$ or when $t_j = n$ and $z_{0j} = 1$, if the estimator (7) is set equal to $\bar{y}$ in these cases. To set the estimator $\phi(x_0)$ equal to $\bar{y}$ when $t_j = 0, n$ is one of the three possibilities given in (3.3.1(15)). The identity is not true when $t_j = 0$ and $z_{0j} = 1$, or when $t_j = n$ and $z_{0j} = 0$, because the estimator (2) is then set equal to 0 and the estimator (7) is still set equal to $\bar{y}$.

4.2.5 ψ and ω procedures

Certain simplifications of the two procedures ψ and ω of Chapter 3.4.2 can be made in the Bernoulli case. If we start with the estimator ψ_r defined in (3.4.2(4)) then it is possible to change the order of summation and due to the theory of sampling distributions we obtain

$$\begin{aligned}
 \psi_r &= \frac{1}{\binom{p}{r}} \sum_{|A|=r} \frac{1}{\binom{p}{r}} \sum_{|B|=r} \phi_B(x_0) = \left(\frac{1}{\binom{p}{r}}\right)^2 \sum_{|B|=r} \sum_{|A|=r} \phi_B(x_0) = \\
 &= \left(\frac{1}{\binom{p}{r}}\right)^2 \sum_{|B|=r} \sum_{|A|=r} (a_B + b_B v(\frac{z}{z_0}(\bar{A}))) = \frac{1}{\binom{p}{r}} \sum_{|B|=r} (a_B + b_B s/p) = \quad (1) \\
 &= \sum_{|B|=r} a_B / \binom{p}{r} + s/p \sum_{|B|=r} b_B / \binom{p}{r}
 \end{aligned}$$

because

$$\frac{1}{\binom{p}{r}} \sum_{|A|=r} v(\frac{z}{z_0}(\bar{A})) = s/p \quad (2)$$

where s denotes the sum of the current data $\underline{z}_0$.

At the calculation of the $\binom{p}{r}$ coefficients a and b we use the same simplification mentioned in (4.2.3) as for the ϕ procedure. The estimator ϕ_r has further, in the Bernoulli case, certain similarities with the estimator ψ_r in the way that the $\binom{p}{r}$ different a_B are combined in the same way as in (1) but the different b_B are weighed with $v(\underline{z}_0^{(B)}) = x_0$ in the estimator ϕ_r .

In order to obtain, in the Bernoulli case, the estimator ω_r defined in (3.4.2(8)) it is possible to make use of the principle which has been suggested in Chapter 4.2.2. The estimator ω_r can be written as

$$\omega_r = \begin{cases} a_r + b_r \frac{1}{\binom{p}{r}} \sum_{\substack{A \\ |A|=r}} v(\underline{z}_0^{(A)}) = a_r + b_r s/p & \text{if not all } s_i = 0, p \\ s/p & \text{if all } s_i = 0, p \end{cases} \quad (3)$$

where s/p is defined in the expression (2) and where s_i denotes the sum of the i th sample $\underline{z}_i$. The coefficients a_r and b_r can be expressed in terms of x and y defined in (3.4.2(6) and (7)) as

$$a_r = \bar{y} - b_r \bar{x} = \frac{\sum_{i=1}^n \sum_B u(\underline{z}_i^{(B)})}{n \binom{p}{r}} - b_r \frac{\sum_{i=1}^n \sum_B u(\underline{z}_i^{(B)})}{n \binom{p}{r}} =$$

$$= \frac{\sum_{i=1}^n s_i}{np} - b_r \frac{\sum_{i=1}^n s_i}{np} \quad (4)$$

and

$$\begin{aligned}
 b_r &= \frac{\sum \sum x y - \frac{\sum x \sum y}{n_r^{(p)}}}{\sum x^2 - \frac{(\sum x)^2}{n_r^{(p)}}} = \\
 &= \frac{\sum_{i=1}^n \sum_{B, |B|=r} v(\underline{z}_i^{(\bar{B})}) u(\underline{z}_i^{(B)}) - \frac{\sum_{i=1}^n \sum_{B, |B|=r} v(\underline{z}_i^{(\bar{B})}) \sum_{i=1}^n \sum_{B, |B|=r} u(\underline{z}_i^{(B)})}{n_r^{(p)}}}{\sum_{i=1}^n \sum_{B, |B|=r} v^2(\underline{z}_i^{(\bar{B})}) - \frac{(\sum_{i=1}^n \sum_{B, |B|=r} v(\underline{z}_i^{(\bar{B})}))^2}{n_r^{(p)}}} .
 \end{aligned} \tag{5}$$

It is possible, in the Bernoulli case, to express the sums, which b_r is built up of, in terms of s_i . If we start with the first sum of the denominator of b_r in (5) it can be expressed as

$$\sum_{i=1}^n \sum_{B, |B|=r} v^2(\underline{z}_i^{(\bar{B})}) = \sum_{i=1}^n \binom{p}{r} \left(\frac{(p-r)s_i}{rp(p-1)} + \frac{(r-1)s_i^2}{rp(p-1)} \right) . \tag{6}$$

In order to obtain the identity (6) the following expressions have been used

$$\begin{aligned}
 &\sum_{i=1}^n \sum_{B, |B|=r} (v(\underline{z}_i^{(\bar{B})}) - \frac{\sum_{B, |B|=r} v(\underline{z}_i^{(\bar{B})})}{\binom{p}{r}})^2 = \\
 &= \sum_{i=1}^n \left(\sum_{B, |B|=r} v^2(\underline{z}_i^{(\bar{B})}) - \frac{(\sum_{B, |B|=r} v(\underline{z}_i^{(\bar{B})}))^2}{\binom{p}{r}} \right) = \sum_{i=1}^n \binom{p}{r} \sigma^2(v(\underline{z}_i^{(\bar{B})})) .
 \end{aligned} \tag{7}$$

The expression $\sigma^2(v(\underline{z}_i^{(\bar{B})}))$ is the variance of a mean which is built up from r observations drawn randomly without replacement from a two-point distributed population of p elements. Thus the variance is

$$\sigma^2(v(\underline{z}_i^{(\bar{B})})) = \frac{1}{r} \frac{s_i}{p} \left(1 - \frac{s_i}{p}\right) \frac{p-r}{p-1}. \quad (8)$$

By using

$$\sum_{\substack{B \\ |B|=r}} v(\underline{z}_i^{(\bar{B})}) / \binom{p}{r} = \frac{s_i}{p} \quad (9)$$

and the expressions (7) and (8) the desired sum can be obtained as

$$\begin{aligned} \sum_{i=1}^n \sum_{\substack{B \\ |B|=r}} v^2(\underline{z}_i^{(\bar{B})}) &= \sum_{i=1}^n \binom{p}{r} \left(\frac{1}{r} \frac{p-r}{p-1} \left(\frac{s_i}{p} - \frac{s_i^2}{p} \right) + \frac{s_i^2}{p} \right) = \\ &= \sum_{i=1}^n \binom{p}{r} \left(\frac{p-r}{r(p-1)} \frac{s_i}{p} + \frac{s_i^2(rp-r-p+r)}{p^2 r(p-1)} \right) = \\ &= \sum_{i=1}^n \binom{p}{r} \left(\frac{(p-r)s_i}{rp(p-1)} + \frac{s_i^2(r-1)}{rp(p-1)} \right) \end{aligned} \quad (10)$$

which corresponds to the expression (6).

The product sum of b_r in (5) can be reduced to

$$\sum_{i=1}^n \sum_{\substack{B \\ |B|=r}} v(\underline{z}_i^{(\bar{B})}) u(\underline{z}_i^{(B)}) = \sum_{i=1}^n \binom{p}{r} \frac{s_i(s_i-1)}{p(p-1)}. \quad (11)$$

By using the theory of sampling distributions the following can be obtained

$$\sum_{i=1}^n \sum_{\substack{B \\ |B|=r}} v(\underline{z}_i^{(\bar{B})}) u(\underline{z}_i^{(B)}) = \sum_{i=1}^n \binom{p}{r} E v(\underline{z}_i^{(\bar{B})}) u(\underline{z}_i^{(B)}) \quad (12)$$

where

$$\begin{aligned} E v(\underline{z}_i^{(\bar{B})}) u(\underline{z}_i^{(B)}) &= E \frac{\sum_{j=1}^r z_{ij} (s_i - \sum_{j=1}^r z_{ij})}{r(p-r)} = \\ &= \frac{1}{r(p-r)} (s_i E \sum_{j=1}^r z_{ij} - E (\sum_{j=1}^r z_{ij})^2) = \\ &= \frac{1}{r(p-r)} \left(\frac{s_i r s_i}{p} - \frac{r^2 (p-r) s_i}{rp(p-1)} - \frac{r^2 (r-1) s_i^2}{rp(p-1)} \right) = \\ &= \frac{1}{p-r} \left(\frac{(p-1)s_i^2 - (r-1)s_i^2 - (p-r)s_i}{p(p-1)} \right) = \\ &= \frac{s_i(s_i-1)}{p(p-1)}. \end{aligned} \quad (13)$$

In the expression (13) we have used the result established in (6) since the following is valid

$$E \left(\frac{\sum_{j=1}^r z_{ij}}{r} \right)^2 = \sum_{\substack{B \\ |B|=r}} v^2(\underline{z}_i^{(\bar{B})}) / \binom{p}{r}. \quad (14)$$

If we apply the result of deduction (13) to expression (12), we obtain the simplified product sum (11).

If we use the following identities

$$\frac{\sum_B v(\underline{z}_i^{(\bar{B})})}{|B|=r \binom{p}{r}} = \frac{\sum_B u(\underline{z}_i^{(B)})}{|B|=r \binom{p}{p-r}} = \frac{s_i}{p} \quad (15)$$

the two remaining sum expressions of b_r in (5) can be simplified to

$$\frac{\sum_{i=1}^n \sum_B v(\underline{z}_i^{(\bar{B})})}{|B|=r \binom{p}{r}} = \frac{\sum_{i=1}^n \sum_B u(\underline{z}_i^{(B)})}{|B|=r \binom{p}{p-r}} = \quad (16)$$

$$= \frac{(\sum_{i=1}^n \sum_B v(\underline{z}_i^{(\bar{B})}))^2}{n \binom{p}{r}} = \binom{p}{r} \frac{(\sum_{i=1}^n s_i)^2}{np^2}.$$

The direction coefficient, b_r , in (5) can now be written as

$$b_r = \frac{\sum_{i=1}^n \sum_B v(\underline{z}_i^{(\bar{B})}) u(\underline{z}_i^{(B)}) - \frac{\sum_{i=1}^n \sum_B v(\underline{z}_i^{(\bar{B})}) \sum_{i=1}^n \sum_B u(\underline{z}_i^{(B)})}{|B|=r \binom{p}{r}}}{\sum_{i=1}^n \sum_B v^2(\underline{z}_i^{(\bar{B})}) - \frac{(\sum_{i=1}^n \sum_B v(\underline{z}_i^{(\bar{B})}))^2}{n \binom{p}{r}}} = \quad (17)$$

(17) cont.

(17) cont.

$$\begin{aligned}
&= \frac{\sum_{i=1}^n \binom{p}{r} \frac{s_i(s_i - 1)}{p(p-1)} - \binom{p}{r} \frac{\sum_{i=1}^n s_i^2}{np}}{\sum_{i=1}^n \binom{p}{r} \frac{(p-r)s_i + (r-1)s_i^2}{rp(p-1)} - \binom{p}{r} \frac{\sum_{i=1}^n s_i^2}{np}} = \\
&= \frac{\frac{\sum_{i=1}^n s_i(s_i - 1)}{p-1} - \frac{\sum_{i=1}^n s_i^2}{np}}{\frac{(p-r) \sum_{i=1}^n s_i + (r-1) \sum_{i=1}^n s_i^2}{r(p-1)} - \frac{\sum_{i=1}^n s_i^2}{np}}.
\end{aligned}$$

In order to calculate the estimator ω_r in (3) there is no need for knowing the complete samples $\underline{z}_0, \underline{z}_1, \dots, \underline{z}_n$. It is enough to know the sum of the observations of every sample $s, s_1, \dots, s_n$. This is valid also for the estimator ϵ_r defined, for the Bernoulli case, in (4.2.2(12)). Another similarity between these estimators is that they are identical when r equals one.

The estimator ω_1 can be expressed in the following way

$$\omega_1 = \frac{\sum_{i=1}^n s_i}{np} (1 - b_1) + \frac{s}{p} b_1 = \frac{\sum_{i=1}^n s_i}{np} \left(1 - \frac{\frac{\sum_{i=1}^n s_i(s_i - 1)}{p-1} - \frac{\sum_{i=1}^n s_i^2}{np}}{\frac{(p-1) \sum_{i=1}^n s_i + (r-1) \sum_{i=1}^n s_i^2}{r(p-1)} - \frac{\sum_{i=1}^n s_i^2}{np}} \right) + \quad (18)$$

$$+ \frac{s}{p} \frac{\frac{\sum_{i=1}^n s_i(s_i - 1)}{p-1} - \frac{\sum_{i=1}^n s_i^2}{np}}{\frac{(p-1) \sum_{i=1}^n s_i + (r-1) \sum_{i=1}^n s_i^2}{r(p-1)} - \frac{\sum_{i=1}^n s_i^2}{np}} =$$

cont.

$$\begin{aligned}
&= \frac{\sum_{i=1}^n s_i}{np} \frac{\sum_{i=1}^n s_i - \frac{(\sum_{i=1}^n s_i)^2}{np} - \frac{\sum_{i=1}^n s_i(s_i - 1)}{p-1} + \frac{(\sum_{i=1}^n s_i)^2}{np}}{\sum_{i=1}^n s_i - \frac{(\sum_{i=1}^n s_i)^2}{np}} + \\
&+ \frac{s}{p(p-1)} \frac{np \sum_{i=1}^n s_i(s_i - 1) - (p-1)(\sum_{i=1}^n s_i)^2}{np \sum_{i=1}^n s_i - (\sum_{i=1}^n s_i)^2} = \\
&= \frac{1}{p-1} \frac{(p-1) \sum_{i=1}^n s_i - \sum_{i=1}^n s_i^2 + \sum_{i=1}^n s_i}{np - \sum_{i=1}^n s_i} + \\
&+ \frac{s}{p(p-1)} \frac{np \sum_{i=1}^n s_i(s_i - 1) - (p-1)(\sum_{i=1}^n s_i)^2}{np \sum_{i=1}^n s_i - (\sum_{i=1}^n s_i)^2} = \\
&= \frac{\sum_{i=1}^n s_i(p - s_i)}{(p-1) \sum_{i=1}^n (p - s_i)} + \frac{s}{p} \frac{np \sum_{i=1}^n s_i(s_i - 1) - (p-1)(\sum_{i=1}^n s_i)^2}{(p-1)(np \sum_{i=1}^n s_i - (\sum_{i=1}^n s_i)^2)}
\end{aligned}$$

and the estimator ϵ_1 can be expressed in the following way

$$\begin{aligned}
\epsilon_1 &= \frac{1}{p} \left((p-s) \frac{\sum_{i=1}^n s_i(p-s_i)}{(p-1) \sum_{i=1}^n (p-s_i)} + s \frac{\sum_{i=1}^n s_i(s_i-1)}{(p-1) \sum_{i=1}^n s_i} \right) = \frac{\sum_{i=1}^n s_i(p-s_i)}{(p-1) \sum_{i=1}^n (p-s_i)} + \\
&+ \frac{s}{p(p-1)} \frac{(\sum_{i=1}^n s_i^2 - \sum_{i=1}^n s_i)(np - \sum_{i=1}^n s_i) - \sum_{i=1}^n s_i(p \sum_{i=1}^n s_i - \sum_{i=1}^n s_i^2)}{(np - \sum_{i=1}^n s_i) \sum_{i=1}^n s_i} = \\
&= \frac{\sum_{i=1}^n s_i(p-s_i)}{(p-1) \sum_{i=1}^n (p-s_i)} + \frac{s}{p(p-1)} \left\{ np \sum_{i=1}^n s_i^2 - np \sum_{i=1}^n s_i - \sum_{i=1}^n s_i \sum_{i=1}^n s_i^2 + \right. \\
&+ \left(\sum_{i=1}^n s_i \right)^2 - p \left(\sum_{i=1}^n s_i \right)^2 + \sum_{i=1}^n s_i \sum_{i=1}^n s_i^2 \} / \{ np \sum_{i=1}^n s_i - \left(\sum_{i=1}^n s_i \right)^2 \} = \\
&= \frac{\sum_{i=1}^n s_i(p-s_i)}{(p-1) \sum_{i=1}^n (p-s_i)} + \frac{s}{p} \frac{np \sum_{i=1}^n s_i(s_i-1) - (p-1) \left(\sum_{i=1}^n s_i \right)^2}{(p-1) \left(np \sum_{i=1}^n s_i - \left(\sum_{i=1}^n s_i \right)^2 \right)}.
\end{aligned} \tag{19}$$

We can see that the two estimators are identical when r equals one also in the special case when all s_i are equal to 0 or p . If all $s_i = 0$ the estimators ω_1 and ϵ_1 are equal to 0 due to the convention which says that $0/0$ equals 0. If all $s_i = p$ the estimator ω_1 in (3) is defined as s/p , the formula (18) for the estimator ϵ_1 is however equal to 0 due to the convention above, but the formula for the estimator ϵ_1 in (4.2.2(4)) is on the other hand equal to s/p in this case.

5. MONTE CARLO SIMULATION IN THE BERNOULLI CASE

In order to get an idea of the accuracy of the EB estimators in the Bernoulli case a Monte Carlo simulation has been performed. In addition to the EB estimators which have been presented in this thesis a lot of EB estimators taken from the literature will be computed in the Monte Carlo simulation.

5.1 Design

To the i th experiment a value θ_i is generated from the prior distribution, $G(\theta)$. Then, to every experiment, a sample of p observations $z_{i1}, \dots, z_{ip}$ is generated from the distribution $\text{Bernoulli}(\theta_i)$, for $i = 0, 1, \dots, n$. The number of experiments is thus $n + 1$. The experiment 0 is called the current experiment and its parameter value, θ_0 , is to be estimated. The following data matrix can be obtained

$$\begin{array}{c|cccc}
 z_{01} & z_{11} & z_{21} & \dots & z_{n1} \\
 \vdots & \vdots & \vdots & & \vdots \\
 z_{0p} & z_{1p} & z_{2p} & \dots & z_{np} \\
 s_0 & s_1 & s_2 & \dots & s_n \\
 \hline
 \theta_0 & \theta_1 & \theta_2 & \dots & \theta_n
 \end{array} \tag{1}$$

where $s_0, s_1, \dots, s_n$ are the sums of the p observations of every sample. From the data the different estimators, $\pi(\underline{z}_0)$, of the simulation are then computed.

The number of iterations will be $k = 200$. At every iteration a data matrix (1) will be generated, which is the basis for the computations of the different estimators. After 200 iterations there are thus 200 outcomes for each and all of the estimators. For these outcomes the following ratio is computed, for every estimator,

$$R_{n+1} = \frac{\sum_{j=1}^k (\pi(\underline{z}_{0j}) - \theta_{0j})^2 / k}{\sum_{j=1}^k \sum_{i=0}^n (s_{ij}/p - \theta_{ij})^2 / (k(n+1))} \quad (2)$$

which is an estimator of the ratio

$$\frac{\text{average squared error of } \pi(\underline{z}_0)}{\text{average squared error of the ML estimator}} \quad (3)$$

In Chapter 2.3 it was pointed out that the ratio (3) is influenced by the prior distribution nearly entirely through the factor

$$S = \frac{E \text{Var}(s/p | \theta)}{\text{Var } \theta} = \frac{(E \theta - E \theta^2)/p}{\text{Var } \theta} \quad (4)$$

and not so much by the form of the prior distribution. Therefore only two different prior distributions will be used in the simulation. The first, Prior 1 say, is the Beta distribution $\beta(e, f)$ with $E \theta = .2$. The second, Prior 2 say, is of the U type with only three possible outcomes. The Prior 2 is thus a discrete distribution. The expected value of Prior 2 has also been put to .2, thus $E \theta = .2$. Prior 2 is of the form $P(\theta = .2 - x) = P(\theta = .2 + x) = .45$ and $P(\theta = .2) = .1$, where x can be calculated from the value of $\text{Var } \theta$.

Two different values of the factor S are to be used in the simulation. We set $S = 1, 3$.

Three different sizes of the samples are to be used namely $p = 5, 10, 15$. In addition three different number of past samples will be used in the simulation, namely $n = 5, 10, 20$.

The variance of the prior distribution, $\text{Var } \theta$, can be computed from the values given on $E \theta$, S and p . $\text{Var } \theta$ is equal to $2/75, 4/275, 1/100$ when p is equal to 5, 10, 15 respectively and the factor S equals 1. The corresponding values of $\text{Var } \theta$ is equal to $1/100, 4/755, 4/1150$ if the factor S equals 3.

5.2 Procedures introduced in the thesis

The procedures which have been presented in the thesis and which will be used in the simulation are

$$\gamma_r \text{ defined in (3.2.1(39)) and } \gamma_r^{(c)} \text{ defined in (3.2.1(40))} \quad (1)$$

$$\epsilon_r \text{ defined in (4.2.2(12))} \quad (2)$$

$$\phi_r \text{ defined in (3.4.1(20)). Three estimators of this type are} \quad (3)$$

obtained, because the three alternatives defined in (3.4.1(15))
are used when $S_{xx} = 0$

$$\psi_r \text{ defined in (4.2.5(1))} \quad (4)$$

$$\omega_r \text{ defined in (4.2.5(3)).} \quad (5)$$

The estimator ω_r is modified so that two alternatives are obtained. In the first alternative the direction coefficient, b_r , is set equal to 0 if $b_r < 0$ and in the second alternative the direction coefficient, b_r , is set equal to 1 if $b_r < 0$. The estimates of ω_r are in this way never negative. The modification means that if $b_r < 0$ then ω_r equals $\sum_{i=1}^n s_i / np$ in the first alternative and equals s/p in the second.

5.2.1 Choice of r-value and reduction of the number of procedures

To every type of procedure in Chapter 5.2, $p-1$ different estimators can be constructed, because r can take the values $1, \dots, p-1$. A question, which has been asked, is if there is some optimal r -value, i.e. if the corresponding estimator has the least average squared error. The simulation indicates, however, that there does not seem to be any optimal r -value. For one procedure the average squared error can e.g. be least for small r -values when S equals three while the tendency can be the opposite when S equals one, if it for the rest is the same constellation of Prior, n and p .

What can be done to get only one estimate from every procedure? Here we have four suggestions:

1. The average of the $p-1$ estimates is used.
2. The median of the $p-1$ estimates is used.
3. The outcome s of the current sample is used in such a way that the $\left[\frac{(p-1)s}{2m} + .5\right]$ th estimate in order of magnitude of the $p-1$ estimates is chosen, where $[\cdot]$ denotes the integer part and where $m = \sum_{i=1}^n s_i/n$. If $[\cdot] < 1$ then $[\cdot] = 1$ and if $[\cdot] > p-1$ then $[\cdot] = p-1$.
4. As 3., but s is here compared with the median, md , of the outcomes $s_1, \dots, s_n$. If $s < md$ then the least estimate is chosen. If $s > md$ then the biggest estimate is chosen. If $s = md$ then the average of the least and the biggest estimate is used.

The four suggestions above will be used in the simulation for every procedure (5.2(1)–(5)).

Owing to the cost the number of procedures, which the simulation comprises, has to be reduced. Therefore when $p = 15$ only the procedure ω_r defined in (4.2.5(3)) plus the comparison procedures defined in Chapter 5.3 will be computed. The reason why the procedure ω_r is chosen among the procedures (5.2(1)–(5)) is because it usually has a low estimated average squared error in the simulation for $p = 5, 10$ and because the estimates of the procedure ω_r are very quickly computed in the Bernoulli case when the simplifications which have been done in Chapter 4.2.5 are made.

5.3 Comparison procedures

In order to compare the estimation procedures which have been presented in this thesis with procedures, which have been presented earlier in the literature, we will also compute the estimators of eight comparison procedures and the estimators defined in Chapter 5.2 and 5.2.1 in the simulation. One of the eight is the maximum likelihood (ML) estimator. Every procedure is compared with the ML estimator by means of the ratio R_{n+1} defined in (5.4(2)). One of the eight is non-parametric. In the case of the others which require the conditional distribution, $f(z|\theta)$, to be known, the conditional distribution, $f(z|\theta)$, is assumed on one hand to be a binomial distribution and on the other to be a

Poisson distribution. In that way two estimators of every procedure are obtained and an indication about the robustness of the procedures is given. The robustness which is indicated is the robustness against an incorrect assumption about the conditional distribution, $f(z|\theta)$. This distribution assumption must be made when the estimators are to be constructed. The binomial distribution and the Poisson distribution agree very closely with each other for small θ values. Thus, we cannot expect any large differences between the estimators.

The eight comparison procedures are:

1. The ML estimator

$$m = s/p \quad (1)$$

2. The Robbins' estimator, Robbins (1955), in the binomial case

$$\pi(s) = \frac{s+1}{p} \frac{f_{n,p}(s+1)}{f_{n,p-1}(s)} \quad (2)$$

where f is defined in connection to (2.1(18)).

The Robbins' estimator in the Poisson case

$$\pi(s) = \frac{s+1}{p} \frac{f_n(s+1)}{1 + f_n(s)} \quad (3)$$

where f is defined in connection to (2.1(9)).

3. The Parametric estimator, Maritz (1970), in the binomial-Beta case

$$\pi(s) = \frac{\hat{e} + s}{p + \hat{e} + \hat{f}} \quad (4)$$

where (4) has been presented in (2.1(41)) and where $\hat{e}$ and $\hat{f}$ are defined in (2.1(42) and (43)) if $z_i = s_i$.

The Parametric estimator in the Poisson-Gamma case

$$\pi(s) = \begin{cases} \frac{\hat{\beta} + s}{p(\hat{\alpha} + 1)} & \text{if } (m_2 - (n+1)m_1^2)/n - m_1 \leq 0 \\ m_1/p & \text{otherwise} \end{cases} \quad (5)$$

where

$$\hat{\alpha} = m_1/\hat{\beta} \quad (6)$$

$$\hat{\beta} = m_1^2/((m_2 - (n+1)m_1^2)/n - m_1) \quad (7)$$

and where

$$m_1 = \frac{\sum_{i=0}^n s_i}{n+1} \quad (8)$$

$$m_2 = \frac{\sum_{i=0}^n s_i^2}{n+1} \quad (9)$$

The estimator (5) is from Maritz (1970, page 52) with (6) and (7) from Maritz (1970, page 41).

4. The Ergle estimator (non-parametric), Ergle and Krutchkoff (1972)

$$\lambda(s) = \frac{\sum_{i=0}^n \frac{s_i}{p} \hat{f}(\underline{z}_i | \theta_i)}{\sum_{i=0}^n \hat{f}(\underline{z}_i | \theta_i)} \quad (10)$$

where

$$\left\{ \begin{aligned} \hat{f}(\underline{z}|\theta_i) &= \prod_{j=1}^p \hat{f}(z_j|\theta_i) \\ \hat{f}(z_j|\theta_i) &= 1/(2\pi p^{4/5}) \sum_{k=1}^p \left(\frac{\sin[A_{ij}(z_k)]}{A_{ij}(z_k)} \right)^2 \\ A_{ij}(z_k) &= (z_k - z_{ij})/(2h_i) \\ h_i &= \left(\frac{1}{p}\right)^{7/10} \left(\sum_{j=1}^p (z_{ij} - \frac{s_i}{p})^2 \right)^{1/2} \end{aligned} \right. \quad (11)$$

If $h_i = 0$ we put $f(z|\theta) = 0$. Besides we define $\sin 0/0 \equiv 1$. The estimator (10) is defined in (2.2(25)) for the general case.

5. The Lemon estimator, Lemon and Krutchkoff (1969), in the binomial case

$$\pi(s) = \left\{ \begin{aligned} &0 && \text{if all } s_i = 0 \\ &\frac{\sum_{i=0}^n \left(\frac{s_i}{p}\right)^{s+1} \left(1 - \frac{s_i}{p}\right)^{p-s}}{\sum_{i=0}^n \left(\frac{s_i}{p}\right)^s \left(1 - \frac{s_i}{p}\right)^{p-s}} && \text{if not all } s_i = 0, p \\ &1 && \text{if all } s_i = p \end{aligned} \right. \quad (12)$$

The estimator (12) is Robbins' original estimator (2.1(16)) if $G(\theta)$ is set equal to its empirical distribution function.

The authors suggest an iteration of (12). In the binomial case the iteration estimator is equal to

$$\pi(s) = \begin{cases} 0 & \text{if all } \pi(s_i) = 0 \\ \frac{\sum_{i=0}^n \{\pi(s_i)\}^{s+1} \{1 - \pi(s_i)\}^{p-s}}{\sum_{i=0}^n \{\pi(s_i)\}^s \{1 - \pi(s_i)\}^{p-s}} & \text{if not all } \pi(s_i) = 0, 1 \\ 1 & \text{if all } \pi(s_i) = 1 \end{cases} \quad (13)$$

where $\pi(s_i)$ are obtained from (12).

In the Poisson case the Lemon estimator is

$$\pi(s) = \begin{cases} 0 & \text{if all } s_i = 0 \\ \frac{\sum_{i=0}^n e^{-s_i} s_i^{s+1}}{p \sum_{i=0}^n e^{-s_i} s_i^s} & \text{otherwise} \end{cases} \quad (14)$$

and the iteration estimator is equal to

$$\pi(s) = \begin{cases} 0 & \text{if all } \pi(s_i) = 0 \\ \frac{\sum_{i=0}^n \exp\{-\pi(s_i)\} \{\pi(s_i)\}^{s+1}}{p \sum_{i=0}^n \exp\{-\pi(s_i)\} \{\pi(s_i)\}^s} & \text{otherwise} \end{cases} \quad (15)$$

where $\pi(s_i)$ are obtained from (14).

The estimators (14) and (15) have been defined in (2.1(24) and (22)).

6. The Griffin estimator a, Griffin and Krutchkoff (1971 a)

$$\pi(s) = \begin{cases} 0 & \text{if all } s_i = 0 \\ \frac{\sum_{i=0}^n \left(\frac{s_i}{p}\right)^{s+1} \left(1 - \frac{s_i}{p}\right)^{p-s-1}}{\sum_{i=0}^n \left(\frac{s_i}{p}\right)^s \left(1 - \frac{s_i}{p}\right)^{p-s-1}} & \text{if not all } s_i = 0, p \\ & \text{or if } s \neq p \\ 1 & \text{if all } s_i = p \text{ or if } s = p \end{cases} \quad (16)$$

If we compare the estimator (16) with the Lemon estimator (12) we can see that the estimator (16) is Robbins' modified estimator defined in (2.1(17)), if the prior distribution, $G(\theta)$, is put equal to its empirical distribution function. The estimator (16) has been defined in (2.1(27)), but there the loss function was defined as $L(\eta, \theta) = \frac{(\eta - \theta)^2}{\theta(1 - \theta)}$, while the loss function here is defined as $L(\eta, \theta) = (\eta - \theta)^2$.

7. The Griffin estimator b, Griffin and Krutchkoff (1971 b), is expressed in the binomial case as

$$\pi(s) = \beta_{n+1} \frac{s}{p} + \alpha_{n+1} \quad (17)$$

where

$$\beta_{n+1} = \begin{cases} 1 & \text{if } (p-1)m_2 + m_1 - pm_1^2 \leq 0 \text{ or if } m_2 - m_1^2 \leq 0 \\ \frac{p(m_2 - m_1^2)}{(p-1)m_2 + m_1 - pm_1^2} & \text{otherwise} \end{cases} \quad (18)$$

$$\alpha_{n+1} = (1 - \beta_{n+1})m_1 \quad (19)$$

and where

$$m_1 = \frac{1}{n+1} \sum_{i=0}^n \frac{s_i}{p} \quad (20)$$

$$m_2 = \frac{1}{n+1} \sum_{i=0}^n \frac{s_i(s_i - 1)}{p(p-1)} . \quad (21)$$

The estimator (17) is defined above in (2.1(39)).

In the Poisson case the parameters β and α defined in (2.1(28) and (29)) can be expressed as

$$\beta = \frac{\text{Var } \theta}{\text{Var } s} = \frac{E \theta^2 - E^2 \theta}{E \theta + E \theta^2 - E^2 \theta} \quad (22)$$

$$\alpha = (1 - \beta) E \theta . \quad (23)$$

Now we have

$$E_{\theta} E(s|\theta) = E \theta \quad (24)$$

$$E_{\theta} \{E(s^2 - s)|\theta\} = E \theta^2 . \quad (25)$$

Thus we estimate β and α with

$$\beta'_{n+1} = \begin{cases} 1 & \text{if } m_1 + m_2 - m_1^2 \leq 0 \text{ or if } m_2 - m_1^2 \leq 0 \\ \frac{m_2 - m_1^2}{m_1 + m_2 - m_1^2} & \text{otherwise} \end{cases} \quad (26)$$

$$\alpha'_{n+1} = (1 - \beta'_{n+1})m_1 \quad (27)$$

where

$$m_1 = \frac{1}{n+1} \sum_{i=0}^n s_i \quad (28)$$

$$m_2 = \frac{1}{n+1} \sum_{i=0}^n s_i (s_i - 1) . \quad (29)$$

The estimator of $\theta = \theta'/p$ is in the Poisson case

$$\pi(s) = (\beta'_{n+1} s + \alpha'_{n+1})/p . \quad (30)$$

8. The Martz estimator, Bennett and Martz (1972), can in the binomial case be expressed as

$$\pi(s) = \frac{\max_{\int} \theta f(s|\theta) g_{n+1}(\theta) d\theta}{\min_{\int} f(s|\theta) g_{n+1}(\theta) d\theta} = \frac{\max_{\int} \theta^{s+1} (1-\theta)^{p-s} g_{n+1}(\theta) d\theta}{\min_{\int} \theta^s (1-\theta)^{p-s} g_{n+1}(\theta) d\theta} \quad (31)$$

where $g_{n+1}(\theta)$ is estimated by means of Parzen's density estimator in the following way

$$\left\{ \begin{aligned} g_{n+1}(\theta) &= 1/(2\pi(n+1)^{4/5}) \sum_{i=0}^n \left\{ \frac{\sin A_i(\theta)}{A_i(\theta)} \right\}^2 \\ A_i(\theta) &= (\theta - \frac{s_i}{p})/(2h) \\ h &= \left(\frac{1}{n+1} \right)^{7/10} \left\{ \sum_{i=0}^n \left(\frac{s_i}{p} - \frac{1}{n+1} \sum_{i=0}^n \frac{s_i}{p} \right)^2 \right\}^{1/2} \end{aligned} \right. \quad (32)$$

The numerical integration is carried out with Romberg's method cf. Lundqvist and Sjöberg (1969) and the limits of integration are set equal to

$$\min = \min_{i=0, \dots, n} \left(\frac{s_i}{p} \right) \quad (33)$$

$$\max = \max_{i=0, \dots, n} \left(\frac{s_i}{p} \right) . \quad (34)$$

In the Poisson case the Martz estimator is equal to

$$\pi(s) = \frac{\max_{\min} \int_{\min}^{\max} \theta f(s|\theta) g_{n+1}(\theta) d\theta}{p \int_{\min}^{\max} f(s|\theta) g_{n+1}(\theta) d\theta} = \frac{\max_{\min} \int_{\min}^{\max} e^{-\theta} \theta^{s+1} g_{n+1}(\theta) d\theta}{p \int_{\min}^{\max} e^{-\theta} \theta^s g_{n+1}(\theta) d\theta} \quad (35)$$

where

$$\left\{ \begin{array}{l} g_{n+1}(\theta) = 1/(2\pi(n+1)^{4/5}) \sum_{i=0}^n \left\{ \frac{\sin A_i(\theta)}{A_i(\theta)} \right\}^2 \\ A_i(\theta) = (\theta - s_i)/(2h) \\ h = \left(\frac{1}{n+1} \right)^{7/10} \left\{ \sum_{i=0}^n \left(s_i - \frac{1}{n+1} \sum_{i=0}^n s_i \right)^2 \right\}^{1/2} \end{array} \right. \quad (36)$$

The limits of integration are here set equal to

$$\min = \min_{i=0, \dots, n} (s_i) \quad (37)$$

$$\max = \max_{i=0, \dots, n} (s_i) . \quad (38)$$

5.4 Result and discussion

The non-parametric procedures γ , ϵ , ϕ , ψ and ω , presented in Chapter 3 and 4, depend on r . In the construction of the non-parametric procedures the samples will be partitioned into two parts, where the parts are composed of r and $p - r$ observations. The parts composed of r observations are used to construct the EB estimator and the parts composed of $p - r$ observations are used to estimate it.

Johns (1957) suggested the estimator λ_p defined in (2.2(5)) which is equivalent to the estimator γ_{p-1} , i.e. he sets r equal to $p - 1$. Maritz (1970) modified λ_p by permuting the past data and obtained an estimator which is equivalent to the estimator ϵ_{p-1} . Neither Johns nor Maritz thought of the possibility of varying the r value. They used only one observation to estimate the Bayes estimator.

No investigation of the accuracy of the estimators of Johns and Maritz has been worked out before, except for Johns' calculation of upper and lower bounds of the extra risk of the estimator λ_p , as has been mentioned in Chapter 2.2. Johns has however shown that the estimator λ_p is asymptotically optimal.

The procedures ϕ , ψ and ω are constructed with the aid of the regression method. Krutchkoff (1967) has constructed a non-parametric EB estimator with the regression method in a situation where, besides the past data, there is assumed to be supplementary estimates of the past parameters available. When θ_0 is to be estimated we have thus the data $x_0, x_1, \dots, x_n$ where x_i is an outcome of the random variable with the conditional distribution $F(x|\theta_i)$, $i = 0, 1, \dots, n$. The supplementary data $y_1, \dots, y_n$ independently obtained from the conditional distribution $H(y|\theta_i)$, $i = 1, \dots, n$ are also available. Krutchkoff constructs firstly one estimator which agrees very closely with Johns' estimator and secondly a regression estimator, which is equivalent to the estimator $\phi(x_0)$, defined in (3.4.1(15), where he defined his estimator as 0 when $S_{xx} = 0$), if the data are obtained in Krutchkoff's way and not as in this work. Krutchkoff shows in the article that his two estimators are asymptotically optimal in the supplementary data situation.

Krutchkoff refers to a simulation he has performed when the problem is to estimate θ . The prior distribution is a normal distribution with $E \theta = .2$ and

$\text{Var } \theta = .0025$. The variable x has the binomial distribution, $\text{bin}(p, \theta)$, with $p = 5$ and the variable y is $\text{bin}(10, \theta)$. In the simulation the author is able to estimate the average squared error of the two non-parametric EB estimators. He lets the number of past experiments, n , vary from 1 to 30. The estimated average squared errors decrease when n increases. The regression estimator has a rather lower estimated average squared error than the other estimator for the n values observed. Krutchkoff declares that both estimators have an average squared error smaller than that of the ML estimator, $x/5$, for as little as three past experiments. We must remember, however, that in the situation observed there is a large S value, $S = 12.6$. The S factor is defined in (2.3(4) and 5.1(4)). The more S increases the better are the EB procedures compared with the classical procedures.

No simulation which deals with a large number of the 'comparison procedures', defined in Chapter 5.3, at the same time, has been presented in the literature. Simulations in which one procedure is compared with the ML estimator have, however, been undertaken. A lot of articles have been reported in Chapter 2 and there the results of the simulations, presented in the articles, can be found. Here we only give a short summary of the estimators which have been calculated in the simulations and for which conditional distributions the simulations have been performed.

The two 'Lemon', 'Lemon' and 'Lemon it.'*, have been compared with the ML estimator in the Poisson case, Lemon and Krutchkoff (1969). 'Griffin a' and 'Griffin b' have been compared with the ML estimator in the binomial case, Griffin and Krutchkoff (1971 a, b). In the simulation of 'Griffin a', the usual quadratic loss function has not been used. 'Martz' has been compared with the ML estimator in the Poisson case, Bennett and Martz (1972). In that simulation comparisons have also been made with the two 'Lemon' and an estimator constructed by Maritz. The non-parametric 'Ergle' procedure has been compared with the various classical estimators for the mean, the variance, the standard deviation, the mode and the 80th percentile, Ergle and Krutchkoff (1972). In that simulation the conditional distribution was chosen from the Pearson family.

The result of the simulation undertaken by the author is presented in the Appendix in the form of a table for the ratio R_{n+1} for the different estimators. The ratio R_{n+1} is defined in (5.1(2)). In addition, the table gives information

*) Cf. Appendix

about the size of the estimated standard errors of the datapoints.

Firstly we will discuss what happens to R_{n+1} of the different estimators when the different quantities n , S , p and Prior are varied. Secondly we discuss if it is possible from the simulation to find an optimal r . Thirdly we try from the table of the Appendix to compare the non-parametric EB procedures of the thesis γ , ϵ , ϕ , ψ and ω with each other. Fourthly we consider the 'comparison procedures'. Finally we compare the best non-parametric EB estimators with the best 'comparison procedures'.

If, in the table of Appendix, we consider the different quantities which are varied in the simulation, we can see, as expected, that R_{n+1} decreases as n increases. Given n , the ratio R_{n+1} of most of the procedures seems to depend only on the factor S . This has also been established by other authors, cf. Chapter 2.

As has already been pointed out in Chapter 2, Ergle and Krutchkoff (1972) assert that R_{n+1} shows a tendency to increase when p increases. This is not evident from the result of the present simulation, with the exception of 'Ergle', 'Lemon' and 'Griffin a', which seem to depend on p . The dependence will here give the opposite effect to that mentioned by Ergle and Krutchkoff. (See below for further comments.) The ratio R_{n+1} of the other procedures does not seem to be dependent on p . The observation of Ergle and Krutchkoff may depend on the fact that they have perhaps not let $\text{Var } \theta$ increase when p increases, which means that the factor S is not kept constant. When p increases the factor S decreases which results in the fact that R_{n+1} increases.

Between Prior 1 and Prior 2 there does not seem to be any differences with regard to the outcomes of R_{n+1} for different n , S and p . It ought to be pointed out that the regression model presented in Chapter 3.4 is true at Prior 1 but false at Prior 2. This does not seem to influence the outcome of R_{n+1} for the procedures which are constructed by means of the regression procedure. The procedures seem to be robust with regard to deviation from the model. To be able to infer with more confidence, the simulation would of course have to comprise more priors.

In the simulation, $E \theta$ has not been varied. Thus, we do not investigate if the ratio R_{n+1} of the different procedures are influenced by $E \theta$.

If the values of the factor R_{n+1} , which are given by the non-parametric

procedures presented in Chapter 3 and 4 of the thesis, are studied in the table, we cannot discern any tendency to an optimal r value. However, it seems to be the estimators with r about $p/2$ which generally give the least R_{n+1} values for the different non-parametric procedures. It is, however, possible to state that the R_{n+1} values which are given by γ_{p-1} and ϵ_{p-1} are the largest of the R_{n+1} values given by all the $p-1$ γ and ϵ estimators. The estimators γ_{p-1} and ϵ_{p-1} are equivalent to the estimators which have been suggested by Johns (1957) and Maritz (1970). The estimators can evidently be improved by choosing another r value. Nobody seems to have thought about trying to find an optimal r before.

Since no optimal r can be chosen, at least not from this simulation, it is convenient, in the comparison of the procedures presented in Chapter 3 and 4, to study the estimators $\cdot$, mean; $\cdot$, median; $\cdot$, s, m; $\cdot$, s, md for the different procedures. These concepts express the fact that the $p-1$ different estimators which can be constructed for every procedure have been gathered in one estimator, cf. the definitions in (5.2.1 point 1., 2., 3., 4. respectively). It seems that the R_{n+1} values are, broadly speaking, of the same size for all the procedures of Chapter 3 and 4. The procedure ϵ seems, however, to give higher R_{n+1} values than the others. This is also valid, but to a lesser degree, for the procedure γ . On the other hand the ω procedure seems to give, on the whole, low R_{n+1} values, particularly ω , s, m and ω , s, md. This is also valid for the estimators $\gamma^{(c)}$, mean, $\gamma^{(c)}$, median, ϕ'' , mean and ϕ'' , median. The last four estimators mentioned seem, in addition, to give particularly low R_{n+1} values when n is small.

Among the estimators presented in Chapter 3 and 4, one of the estimators $\gamma^{(c)}$, mean, $\gamma^{(c)}$, median, ϕ'' , mean or ϕ'' , median seems to be a good choice. Some improvement seems possible if we switch over to one of the estimators, ω , s, m or ω , s, md when n becomes large. We must remember, however, that no general statements can be stipulated from simulations. In addition, in this simulation we have a small number of constellations of Prior, S, n and p . Since the number of iterations is 200, the variation of certain R_{n+1} values is rather large.

If we consider the 'comparison procedures' we can see that 'Robbins' and 'Robbins, P' give very large R_{n+1} values. It has already been mentioned in Chapter 2 that Maritz (1966) has given an example to show that Robbins' esti-

mator in the Poisson case is worse than conventional estimators, unless n is reasonably large. Thus, this disadvantage of Robbins' estimator seems to be true also in the binomial case. The estimators 'Parametric' and 'Martz' on the other hand give small R_{n+1} values. The R_{n+1} values given by 'Martz' do not seem to depend markedly on the Prior, n , p or on the use of the incorrect assumption that $f(z|\theta)$ is a Poisson distribution. That the ratio R_{n+1} does not diminish when n grows is surprising. 'Parametric' on the other hand seems to depend on n and S but not on Prior, which is notable since 'Parametric' is constructed from the assumption of Prior 1. When $f(z|\theta)$ is assumed to be a Poisson distribution the R_{n+1} values are raised for $S = 1$ and diminished for $S = 3$. I.e., an incorrect assumption about $f(z|\theta)$ seems to affect 'Parametric'.

'Ergle', the non-parametric EB estimator among the 'comparison procedures', seems to perform poorly. If we compare 'Ergle', the different 'Lemon' and 'Griffin a' we can see that the R_{n+1} values, which are given by these procedures, seem to depend on p . When p increases the ratio R_{n+1} decreases. These procedures have the similarity that they all are constructed in such a way that the prior distribution $G(\theta)$ which is a part of the estimators is estimated by its empirical distribution function.

In a comparison of $\gamma^{(c)}$, mean, $\gamma^{(c)}$, median, ϕ'' , mean, ϕ'' , median, ω, s, m and ω, s, md with 'Parametric' and 'Martz', i.e. the estimators which have shown to give the smallest R_{n+1} values of the estimators investigated, we can see that the non-parametric ones come out well. Since the ratio R_{n+1} of 'Martz' does not seem to decrease with increasing n , it is probable, as we can see in the table of Appendix, that the R_{n+1} values of the non-parametric EB estimators mentioned above will stabilize themselves on a lower level than the R_{n+1} values of 'Martz' when n increases. It has already been pointed out that when $f(z|\theta)$ is incorrectly assumed to be a Poisson distribution, 'Parametric' will be influenced but not 'Martz'. Whether this is a hint of the robustness of 'Martz' against incorrect assumptions of $f(z|\theta)$ or not, is difficult to say, since the Poisson distribution and the binomial distribution agree very closely for small θ values.

APPENDIX

Table for the ratio R_{n+1}^* for different estimators and for different constellations of prior distribution, S value, sample size, p, and number of past samples, n. The values of the table have been obtained from Monte Carlo simulations.

Estimator	p	S	Prior 1			Prior 2		
			n			n		
			5	10	20	5	10	20
γ_1	5	1	1.32 ² ₁	1.04 ²	0.75 ¹	1.20 ¹	1.05 ¹	0.78
		3	0.77 ² ₂	0.47 ¹	0.43 ¹	0.64 ¹	0.49 ¹	0.34
	10	1	1.44 ¹	1.03 ¹	0.91 ¹	1.45 ¹	1.18 ¹	0.94
		3	0.62 ¹	0.48	0.36	0.65	0.44	0.40
γ_2	5	1	1.27 ² ₁	1.01 ²	0.77 ¹	1.13 ¹	1.01 ¹	0.62
		3	0.81 ³ ₃	0.52 ¹	0.43 ¹	0.65 ¹	0.58 ¹	0.35
	10	1	1.45 ¹ ₁	0.97 ¹	0.87 ¹	1.31 ¹	1.09 ¹	0.86
		3	0.60 ¹	0.51	0.35	0.68	0.46	0.40
γ_3	5	1	1.39 ² ₁	1.03 ²	0.81 ²	1.20 ¹	1.07 ¹	0.58
		3	0.92 ³ ₃	0.63 ¹	0.47 ¹	0.73 ¹	0.74 ¹	0.37
	10	1	1.51 ³ ₁	0.94 ¹	0.83 ¹	1.29 ¹	1.03 ¹	0.76
		3	0.62 ¹	0.53 ¹	0.34	0.71	0.49	0.39
γ_4	5	1	1.63 ³ ₂	1.35 ³ ₂	1.08 ² ₁	1.70 ² ₂	1.51 ² ₂	0.80 ¹ ₁
		3	1.38 ³ ₃	1.09 ¹	0.77 ¹	1.18 ¹	1.21 ¹	0.61 ¹
	10	1	1.56 ³ ₁	0.94 ¹	0.82 ¹	1.30 ¹	0.98 ¹	0.71
		3	0.63 ¹	0.57 ¹	0.35	0.77 ¹	0.52	0.39
γ_5	10	1	1.64 ³ ₁	0.97 ²	0.83 ¹	1.33 ¹	0.98 ¹	0.67
		3	0.66 ¹	0.61 ¹	0.36	0.84 ¹	0.55	0.41
γ_6	10	1	1.76 ³ ₁	1.02 ²	0.87 ²	1.40 ²	1.00 ¹	0.65
		3	0.69 ¹	0.66 ¹	0.39	0.95 ¹	0.59	0.43
γ_7	10	1	1.88 ³ ₁	1.12 ²	0.98 ²	1.52 ²	1.05 ¹	0.65 ¹
		3	0.73 ¹	0.75 ¹	0.45	1.11 ¹	0.65 ¹	0.46
γ_8	10	1	2.13 ³ ₁	1.29 ²	1.16 ² ₁	1.74 ² ₁	1.22 ² ₁	1.74 ¹
		3	0.83 ¹	0.94 ¹	0.58	1.37 ¹	0.82 ¹	0.55
γ_9	10	1	2.70 ³ ₃	1.65 ³ ₃	1.89 ³ ₁	2.44 ³ ₂	1.74 ² ₂	1.54 ³ ₁
		3	1.35 ¹	1.65 ¹	1.00 ¹	1.95 ²	1.43 ²	0.99 ¹
$\gamma_1^{(c)}$	5	1	0.84 ¹ ₁	0.87 ¹	0.75 ¹	0.85 ¹	0.92	0.77
		3	0.61 ¹	0.43 ¹	0.42 ¹	0.48 ¹	0.42 ¹	0.33
	10	1	0.90 ¹ ₁	0.91 ¹	0.89 ¹	1.09 ¹	1.09 ¹	0.93
		3	0.55 ¹	0.46	0.34	0.51	0.37	0.39

*) Estimated standard errors of the data points are below .06.

1) Estimated standard errors of the data points lie between .06 - .12.

2) Estimated standard errors of the data points lie between .12 - .18.

3) Estimated standard errors of the data points are .18 or more.

γ_r , $r = 1, \dots, p-1$, are defined in (3.2.1(39)).

$\gamma_r^{(c)}$, $r = 1, \dots, p-1$, are defined in (3.2.1(40)).

Estimator	p	S	Prior 1			Prior 2		
			n			n		
			5	10	20	5	10	20
$\gamma_2^{(c)}$	5	1	0.71 ¹	0.66 ¹	0.61 ¹	0.74 ¹	0.83 ¹	0.57
		3	0.67 ¹	0.43	0.35	0.53	0.49	0.31
	10	1	0.81 ¹	0.73 ¹	0.73 ¹	0.90 ¹	0.94 ¹	0.79
		3	0.51	0.42	0.31	0.51	0.35	0.37
$\gamma_3^{(c)}$	5	1	0.81 ¹	0.63 ¹	0.57 ¹	0.95 ¹	0.90 ¹	0.54 ¹
		3	0.78 ¹	0.52	0.42	0.66 ¹	0.63	0.39 ¹
	10	1	0.77 ¹	0.65	0.66 ¹	0.83 ¹	0.85 ¹	0.70
		3	0.53	0.43	0.30	0.53	0.36	0.37
$\gamma_4^{(c)}$	5	1	1.09 ²	0.90 ¹	0.71 ¹	1.40 ²	1.36 ²	0.79 ¹
		3	1.12 ¹	0.96 ¹	0.65 ¹	1.15 ²	1.09 ¹	0.63 ¹
	10	1	0.76 ¹	0.61 ¹	0.61 ¹	0.81 ¹	0.79 ¹	0.64
		3	0.54 ¹	0.46	0.29	0.56	0.40	0.38
$\gamma_5^{(c)}$	10	1	0.78 ¹	0.61 ¹	0.60 ¹	0.81 ¹	0.78 ¹	0.60
		3	0.57 ¹	0.51 ¹	0.31	0.63	0.46	0.41
$\gamma_6^{(c)}$	10	1	0.82 ¹	0.65 ¹	0.62 ¹	0.86 ¹	0.80 ¹	0.56 ¹
		3	0.62 ¹	0.58 ¹	0.33	0.73 ¹	0.53 ¹	0.44 ¹
$\gamma_7^{(c)}$	10	1	0.92 ¹	0.72 ¹	0.68 ¹	0.96 ¹	0.87 ¹	0.55 ¹
		3	0.68 ¹	0.68 ¹	0.39	0.89 ¹	0.61 ¹	0.48 ¹
$\gamma_8^{(c)}$	10	1	1.13 ¹	0.88 ¹	0.87 ¹	1.12 ¹	1.02 ¹	0.63 ¹
		3	0.80 ¹	0.87 ²	0.54 ¹	1.18 ¹	0.79 ¹	0.55 ¹
$\gamma_9^{(c)}$	10	1	1.70 ³	1.31 ²	1.58 ³	1.69 ³	1.70 ³	1.13 ²
		3	1.35 ³	1.56 ³	0.94 ¹	1.91 ³	1.41 ²	0.92 ¹
ϵ_1	5	1	1.22 ²	0.91 ¹	0.75 ¹	1.09 ¹	1.00 ¹	0.77
		3	0.73 ²	0.44 ¹	0.41 ¹	0.62 ¹	0.48 ¹	0.34
	10	1	1.27 ²	0.99 ¹	0.90 ¹	1.27 ¹	1.14 ¹	0.95
		3	0.60 ¹	0.47	0.35	0.60	0.42	0.40
ϵ_2	5	1	1.27 ²	0.84 ¹	0.71 ¹	1.05 ¹	0.94 ¹	0.62
		3	0.85 ²	0.49	0.41 ¹	0.68 ¹	0.56 ¹	0.33
	10	1	1.24 ²	0.92 ¹	0.80 ¹	1.17 ¹	1.03 ¹	0.83
		3	0.62 ¹	0.49	0.33	0.63	0.42	0.39
ϵ_3	5	1	1.56 ³	1.03 ²	0.77 ²	1.32 ²	1.05 ¹	0.57
		3	1.21 ²	0.64 ¹	0.46 ¹	0.88 ¹	0.74 ¹	0.37
	10	1	1.28 ²	0.87 ¹	0.74 ¹	1.15 ¹	0.96 ¹	0.74
		3	0.66 ¹	0.52	0.33	0.68	0.43	0.38
ϵ_4	5	1	2.26 ³	1.60 ³	1.02 ²	2.23 ³	1.63 ²	0.84 ¹
		3	2.80 ³	1.21 ²	0.75 ¹	2.28 ³	1.30 ²	0.60 ¹
	10	1	1.38 ³	0.89 ¹	0.71 ¹	1.19 ¹	0.93 ¹	0.69
		3	0.75 ¹	0.57 ¹	0.33	0.78 ¹	0.46	0.39
ϵ_5	10	1	1.61 ³	0.94 ¹	0.72 ¹	1.34 ²	0.94 ¹	0.66
		3	0.89 ¹	0.63 ¹	0.34	0.94 ¹	0.51	0.41
ϵ_6	10	1	2.07 ³	1.04 ²	0.76 ¹	1.71 ³	0.99 ¹	0.65
		3	1.19 ²	0.73 ¹	0.38	1.22 ¹	0.59	0.44

$\epsilon_r, r = 1, \dots, p-1$, are defined in (4.2.2(12)).

Estimator	p	S	Prior 1			Prior 2		
			n			n		
			5	10	20	5	10	20
ϵ_7	10	1	2.96 ³	1.23 ²	0.88 ¹	2.64 ³	1.16 ²	0.68 ¹
		3	1.89 ³	0.95 ¹	0.46	1.88 ³	0.77 ¹	0.49
ϵ_8	10	1	4.86 ³	1.77 ³	1.25 ²	3.46 ³	1.71 ³	0.81 ¹
		3	3.42 ³	1.56 ³	0.69 ¹	3.06 ³	1.25 ²	0.64 ¹
ϵ_9	10	1	10.41 ³	3.72 ³	2.38 ³	6.16 ³	4.27 ³	1.44 ²
		3	7.52 ³	4.62 ³	1.69 ³	7.34 ³	3.65 ³	1.44 ³
ϕ_1	5	1	1.37 ²	1.07 ²	0.75 ¹	1.41 ¹	1.10 ¹	0.78
		3	0.91 ¹	0.54 ¹	0.44 ¹	0.83 ¹	0.58	0.34
	10	1	1.65 ³	1.01	0.92 ¹	1.66 ²	1.23 ¹	0.94
		3	0.84 ¹	0.53	0.35	0.94 ¹	0.49	0.40
ϕ_2	5	1	1.06 ²	0.73 ¹	0.63 ¹	1.04 ¹	0.88 ¹	0.62
		3	0.92 ²	0.48 ¹	0.39 ¹	0.76 ¹	0.54 ¹	0.31
	10	1	1.21 ²	0.88 ¹	0.77 ¹	1.21 ¹	0.99 ¹	0.81
		3	0.68 ¹	0.49	0.32	0.70	0.41	0.38
ϕ_3	5	1	1.02 ²	0.63 ¹	0.60 ¹	1.03 ¹	0.87 ¹	0.56
		3	0.92 ²	0.49	0.40 ¹	0.77 ¹	0.55 ¹	0.32
	10	1	1.11 ²	0.81 ¹	0.69 ¹	1.07 ¹	0.89 ¹	0.72
		3	0.65 ¹	0.50	0.31	0.62	0.41	0.37
ϕ_4	5	1	1.16 ²	0.62 ¹	0.59 ¹	1.44 ³	1.08 ³	0.53
		3	1.09 ²	0.54 ¹	0.44 ¹	0.94 ¹	0.63 ¹	0.34
	10	1	1.06 ²	0.75 ¹	0.64 ¹	1.00 ¹	0.84 ¹	0.67
		3	0.63 ¹	0.51 ¹	0.31	0.62	0.46 ¹	0.37
ϕ_5	10	1	1.07 ²	0.72 ¹	0.61 ¹	0.97 ¹	0.82 ¹	0.63
		3	0.61 ¹	0.54 ¹	0.31	0.66	0.48 ¹	0.37
ϕ_6	10	1	1.02 ²	0.70 ¹	0.59 ¹	0.94 ¹	0.81 ¹	0.60
		3	0.64 ¹	0.53	0.31	0.72 ¹	0.51 ¹	0.37
ϕ_7	10	1	0.99 ¹	0.70 ¹	0.58 ¹	0.93 ¹	0.80 ¹	0.58
		3	0.66 ¹	0.56 ¹	0.34	0.81 ¹	0.50 ¹	0.38
ϕ_8	10	1	0.98 ¹	0.70 ¹	0.58 ¹	0.96 ¹	0.80 ¹	0.56
		3	0.74 ²	0.57 ¹	0.34	0.88 ¹	0.52 ¹	0.39
ϕ_9	10	1	1.05 ²	0.71 ¹	0.58 ¹	1.02 ²	0.81 ¹	0.55
		3	0.95 ³	0.60 ¹	0.35	0.98 ¹	0.55 ¹	0.40
ϕ'_1	5	1	1.18 ²	0.98 ²	0.75 ¹	1.07 ¹	1.00 ¹	0.78
		3	0.68 ¹	0.45 ¹	0.43 ¹	0.57 ¹	0.46 ¹	0.34
	10	1	1.24 ²	0.98 ¹	0.90	1.29 ¹	1.14 ¹	0.94
		3	0.59 ¹	0.47	0.35	0.58	0.41	0.40
ϕ'_2	5	1	1.04 ²	0.72 ¹	0.63 ¹	0.93 ¹	0.86 ¹	0.62
		3	0.81 ²	0.45 ¹	0.39 ¹	0.65 ¹	0.52 ¹	0.31
	10	1	1.08 ²	0.87 ¹	0.77 ¹	1.09 ¹	0.98 ¹	0.81
		3	0.58 ¹	0.48	0.32	0.58	0.40	0.38

ϕ_r , $r = 1, \dots, p-1$, are defined in (3.4.1(20)), equals 0 when $S_{xx} = 0$.

ϕ'_r , $r = 1, \dots, p-1$, are defined in (3.4.1(20)), equals $\bar{y}$ when $S_{xx} = 0$.

Estimator	p	S	Prior 1			Prior 2		
			n			n		
			5	10	20	5	10	20
ϕ'_3	5	1	0.98 ²	0.63 ¹	0.60 ¹	1.02 ¹	0.87 ¹	0.56
		3	0.90 ²	0.48 ¹	0.40 ¹	0.73 ¹	0.54 ¹	0.32
	10	1	1.02 ²	0.80 ¹	0.69 ¹	1.00 ¹	0.89 ¹	0.72
		3	0.60 ¹	0.50	0.31	0.59	0.40	0.37
ϕ'_4	5	1	1.43 ²	0.62 ¹	0.59 ¹	1.40 ³	1.08 ³	0.53
		3	1.04 ²	0.54 ¹	0.44 ¹	0.92 ¹	0.63 ¹	0.34
	10	1	1.04 ²	0.75 ¹	0.64 ¹	1.00 ¹	0.84 ¹	0.67
		3	0.64 ¹	0.51 ¹	0.31	0.61	0.46 ¹	0.37
ϕ'_5	10	1	1.08 ²	0.72 ¹	0.61 ¹	0.94 ¹	0.82 ¹	0.63
		3	0.59 ¹	0.54 ¹	0.31	0.66	0.48 ¹	0.37
ϕ'_6	10	1	1.02 ²	0.70 ¹	0.59 ¹	0.92 ¹	0.81 ¹	0.60
		3	0.63 ¹	0.53	0.31	0.73 ¹	0.51 ¹	0.37
ϕ'_7	10	1	0.99 ¹	0.70 ¹	0.58 ¹	0.92 ¹	0.80 ¹	0.58
		3	0.66 ¹	0.56 ¹	0.34	0.81 ¹	0.50 ¹	0.38
ϕ'_8	10	1	0.98 ¹	0.70 ¹	0.58 ¹	0.95 ¹	0.80 ¹	0.56
		3	0.74 ²	0.57 ¹	0.34	0.88 ¹	0.52 ¹	0.39
ϕ'_9	10	1	1.05 ²	0.71 ¹	0.58 ¹	1.02 ²	0.81 ¹	0.55
		3	0.95 ³	0.60 ¹	0.34	0.98 ¹	0.55 ¹	0.40
ϕ''_1	5	1	0.70 ¹	0.81 ¹	0.74 ¹	0.71 ¹	0.90 ¹	0.76
		3	0.52 ¹	0.43 ¹	0.42	0.52 ¹	0.41 ¹	0.33
	10	1	0.72 ¹	0.79 ¹	0.88 ¹	0.78 ¹	1.01 ¹	0.92
		3	0.48	0.44	0.33	0.47	0.34	0.39
ϕ''_2	5	1	0.81 ¹	0.70 ¹	0.63 ¹	0.80 ¹	0.86	0.62
		3	0.64 ¹	0.45 ¹	0.39	0.60 ¹	0.50	0.31
	10	1	0.85 ¹	0.84 ¹	0.77 ¹	0.89 ¹	0.97 ¹	0.81
		3	0.52	0.48	0.32	0.51	0.38	0.38
ϕ''_3	5	1	0.92 ¹	0.62 ¹	0.60 ¹	1.02 ¹	0.89 ¹	0.56
		3	0.89 ¹	0.47	0.40	0.71 ¹	0.56	0.32
	10	1	0.93 ¹	0.80 ¹	0.69 ¹	0.93 ¹	0.89 ¹	0.72
		3	0.57 ¹	0.50	0.31	0.56	0.40	0.37
ϕ''_4	5	1	1.08 ²	0.62 ¹	0.59 ¹	1.38 ³	1.08 ³	0.53
		3	1.03 ¹	0.54 ¹	0.44 ¹	0.90 ¹	0.63 ¹	0.34
	10	1	0.99 ¹	0.75 ¹	0.64 ¹	0.95 ¹	0.84 ¹	0.67
		3	0.62 ¹	0.51 ¹	0.31	0.59	0.46 ¹	0.37
ϕ''_5	10	1	1.04 ²	0.72 ¹	0.61 ¹	0.93 ¹	0.82 ¹	0.63
		3	0.60 ¹	0.54 ¹	0.31	0.66	0.48 ¹	0.37
ϕ''_6	10	1	0.99 ²	0.70 ¹	0.59 ¹	0.92 ¹	0.81 ¹	0.60
		3	0.63 ¹	0.53	0.31	0.73 ¹	0.51 ¹	0.37
ϕ''_7	10	1	1.00 ¹	0.70 ¹	0.58 ¹	0.93 ¹	0.80 ¹	0.58
		3	0.66 ¹	0.56 ¹	0.34	0.80 ¹	0.50 ¹	0.38
ϕ''_8	10	1	0.98 ¹	0.70 ¹	0.58 ¹	0.96 ¹	0.80 ¹	0.56
		3	0.74 ²	0.57 ¹	0.34	0.89 ¹	0.52 ¹	0.39

ϕ''_r , $r = 1, \dots, p-1$, are defined in (3.4.1(20)), equals s/p when $S_{xx} = 0$.

Estimator	p	S	Prior 1			Prior 2		
			n			n		
			5	10	20	5	10	20
ϕ_9''	10	1	1.05 ² ₃	0.71 ¹ ₁	0.58 ¹	1.02 ² ₁	0.81 ¹ ₁	0.55
		3	0.95 ² ₃	0.60 ¹ ₁	0.35	0.98 ¹ ₁	0.55 ¹ ₁	0.40
ψ_1	5	1	1.17 ² ₁	0.89 ¹ ₁	0.73 ¹ ₁	1.15 ¹ ₁	0.99 ¹ ₁	0.75
		3	0.77 ² ₂	0.46 ¹ ₁	0.41 ¹ ₁	0.70 ¹ ₁	0.50 ¹ ₁	0.33
	10	1	1.26 ² ₁	0.96 ¹ ₁	0.89 ¹ ₁	1.25 ¹ ₁	1.12 ¹ ₁	0.94
		3	0.61 ¹ ₁	0.47 ¹ ₁	0.35 ¹ ₁	0.61 ¹ ₁	0.41 ¹ ₁	0.39
ψ_2	5	1	0.99 ¹ ₂	0.72 ¹ ₂	0.64 ¹ ₁	0.98 ¹ ₁	0.88 ¹ ₁	0.61
		3	0.90 ² ₂	0.45 ¹ ₁	0.39 ¹ ₁	0.74 ¹ ₁	0.54 ¹ ₁	0.31
	10	1	1.12 ² ₁	0.86 ¹ ₁	0.77 ¹ ₁	1.08 ¹ ₁	0.97 ¹ ₁	0.81
		3	0.60 ¹ ₁	0.48 ¹ ₁	0.33 ¹ ₁	0.61 ¹ ₁	0.40 ¹ ₁	0.38
ψ_3	5	1	0.98 ¹ ₂	0.63 ¹ ₁	0.60 ¹ ₁	0.99 ¹ ₁	0.89 ¹ ₁	0.55
		3	0.85 ² ₂	0.47 ¹ ₁	0.40 ¹ ₁	0.76 ¹ ₁	0.55 ¹ ₁	0.31
	10	1	1.07 ² ₁	0.80 ¹ ₁	0.69 ¹ ₁	1.01 ¹ ₁	0.89 ¹ ₁	0.72
		3	0.61 ¹ ₁	0.49 ¹ ₁	0.31 ¹ ₁	0.60 ¹ ₁	0.40 ¹ ₁	0.37
ψ_4	5	1	1.21 ³ ₂	0.62 ¹ ₂	0.60 ¹ ₂	1.37 ³ ₂	1.12 ³ ₂	0.52
		3	1.07 ² ₂	0.52 ¹ ₁	0.43 ¹ ₁	0.95 ¹ ₁	0.61 ¹ ₁	0.34
	10	1	1.04 ² ₁	0.74 ¹ ₁	0.64 ¹ ₁	0.94 ¹ ₁	0.84 ¹ ₁	0.67
		3	0.62 ¹ ₁	0.51 ¹ ₁	0.31 ¹ ₁	0.64 ¹ ₁	0.45 ¹ ₁	0.37
ψ_5	10	1	1.03 ² ₁	0.71 ¹ ₁	0.61 ¹ ₁	0.93 ¹ ₁	0.81 ¹ ₁	0.62
		3	0.59 ¹ ₁	0.54 ¹ ₁	0.30 ¹ ₁	0.69 ¹ ₁	0.47 ¹ ₁	0.37
ψ_6	10	1	1.01 ¹ ₁	0.70 ¹ ₁	0.59 ¹ ₁	0.91 ¹ ₁	0.79 ¹ ₁	0.59
		3	0.63 ¹ ₁	0.53 ¹ ₁	0.31 ¹ ₁	0.74 ¹ ₁	0.48 ¹ ₁	0.37
ψ_7	10	1	0.99 ¹ ₁	0.69 ¹ ₁	0.58 ¹ ₁	0.91 ¹ ₁	0.79 ¹ ₁	0.57
		3	0.65 ¹ ₁	0.55 ¹ ₁	0.33 ¹ ₁	0.80 ¹ ₁	0.48 ¹ ₁	0.38
ψ_8	10	1	0.98 ¹ ₂	0.69 ¹ ₁	0.58 ¹ ₁	0.93 ¹ ₁	0.78 ¹ ₁	0.56
		3	0.73 ¹ ₂	0.58 ¹ ₁	0.34 ¹ ₁	0.87 ¹ ₁	0.50 ¹ ₁	0.39
ψ_9	10	1	1.02 ¹ ₃	0.69 ¹ ₁	0.58 ¹ ₁	1.02 ² ₁	0.79 ¹ ₁	0.55
		3	0.94 ¹ ₃	0.59 ¹ ₁	0.35 ¹ ₁	0.94 ¹ ₁	0.53 ¹ ₁	0.39
ω_1	5	1	1.08 ¹ ₁	0.88 ¹ ₁	0.74 ¹ ₁	1.00 ¹ ₁	0.98 ¹ ₁	0.77
		3	0.63 ² ₂	0.42 ¹ ₁	0.40 ¹ ₁	0.55 ¹ ₁	0.45 ¹ ₁	0.32
	10	1	1.24 ¹ ₁	0.97 ¹ ₁	0.90 ¹ ₁	1.24 ¹ ₁	1.13 ¹ ₁	0.95
		3	0.58 ² ₂	0.45 ¹ ₁	0.34 ¹ ₁	0.57 ¹ ₁	0.41 ¹ ₁	0.39
	15	1	1.20 ² ₂	1.01 ¹ ₁	0.94 ¹ ₁	1.28 ¹ ₁	1.05 ¹ ₁	0.99
ω_2	5	3	0.50 ¹ ₁	0.51 ¹ ₁	0.40 ¹ ₁	0.52 ¹ ₁	0.38 ¹ ₁	0.35
		1	1.00 ¹ ₁	0.75 ¹ ₁	0.66 ¹ ₁	0.91 ¹ ₁	0.89 ¹ ₁	0.65
	10	3	0.62 ² ₂	0.40 ¹ ₁	0.38 ¹ ₁	0.54 ¹ ₁	0.45 ¹ ₁	0.30
		1	1.15 ¹ ₁	0.89 ¹ ₁	0.79 ¹ ₁	1.14 ¹ ₁	1.02 ¹ ₁	0.84
		3	0.56 ² ₂	0.45 ¹ ₁	0.32 ¹ ₁	0.56 ¹ ₁	0.39 ¹ ₁	0.38
		1	1.12 ² ₂	0.94 ¹ ₁	0.85 ¹ ₁	1.21 ¹ ₁	0.97 ¹ ₁	0.91
		3	0.49 ¹ ₁	0.50 ¹ ₁	0.38 ¹ ₁	0.51 ¹ ₁	0.37 ¹ ₁	0.34

ψ_r , $r = 1, \dots, p-1$, are defined in (4.2.5(1)).

ω_r , $r = 1, \dots, p-1$, are defined in (4.2.5(3)), b_r is set equal to 0 when $b_r < 0$.

Estimator	p	S	Prior 1			Prior 2		
			n			n		
			5	10	20	5	10	20
ω_3	5	1	0.95 ¹	0.68 ¹	0.62 ¹	0.87 ¹	0.85 ¹	0.60
		3	0.61 ¹	0.40 ¹	0.38 ¹	0.54 ¹	0.46 ¹	0.30
	10	1	1.10 ²	0.83 ¹	0.72 ¹	1.07 ¹	0.95 ¹	0.77
		3	0.55 ¹	0.45 ¹	0.31 ¹	0.55 ¹	0.38	0.37
	15	1	1.07 ²	0.88 ¹	0.78 ¹	1.15 ¹	0.91	0.85
		3	0.48	0.49	0.37	0.50	0.36	0.33
ω_4	5	1	0.93 ¹	0.65 ¹	0.59 ¹	0.86 ¹	0.83 ¹	0.56
		3	0.61 ¹	0.40 ¹	0.38 ¹	0.54 ¹	0.48 ¹	0.30
	10	1	1.06 ²	0.78 ¹	0.67 ¹	1.02 ¹	0.90 ¹	0.71
		3	0.54 ¹	0.45 ¹	0.30 ¹	0.54 ¹	0.37	0.36
	15	1	1.03 ¹	0.83 ¹	0.73 ¹	1.11 ¹	0.86	0.80
		3	0.47	0.48	0.36	0.50	0.36	0.32
ω_5	10	1	1.04 ²	0.75 ¹	0.64 ¹	0.91 ¹	0.86 ¹	0.67
		3	0.53 ¹	0.45 ¹	0.29 ¹	0.54 ¹	0.36	0.36
	15	1	1.00 ¹	0.79 ¹	0.69 ¹	1.07 ¹	0.83	0.76
		3	0.46	0.47	0.36	0.49	0.35	0.31
ω_6	10	1	1.02 ²	0.73 ¹	0.62 ¹	0.96 ¹	0.83 ¹	0.64
		3	0.53 ¹	0.45 ¹	0.29 ¹	0.54 ¹	0.36	0.36
	15	1	0.98 ¹	0.77 ¹	0.66 ¹	1.05 ¹	0.80	0.73
		3	0.46	0.47	0.35	0.49	0.35	0.31
ω_7	10	1	1.01 ²	0.71 ¹	0.60 ¹	0.95 ¹	0.81 ¹	0.62
		3	0.52 ¹	0.45 ¹	0.29 ¹	0.54 ¹	0.36	0.36
	15	1	0.96 ¹	0.74 ¹	0.63 ¹	1.03 ¹	0.78	0.71
		3	0.46	0.47	0.35	0.49	0.35	0.31
ω_8	10	1	1.00 ²	0.70 ¹	0.59 ¹	0.93 ¹	0.79 ¹	0.60
		3	0.52 ¹	0.46 ¹	0.29	0.54 ¹	0.36	0.36
	15	1	0.95 ¹	0.72 ¹	0.61	1.01 ¹	0.76	0.69
		3	0.46	0.46	0.34	0.49	0.35	0.31
ω_9	10	1	0.99 ²	0.68 ¹	0.58 ¹	0.92 ¹	0.78 ¹	0.58
		3	0.52 ¹	0.46 ¹	0.29	0.54 ¹	0.36	0.36
	15	1	0.94 ¹	0.71 ¹	0.59	1.00 ¹	0.74	0.68
		3	0.45	0.46	0.34	0.49	0.35	0.30
ω_{10}	15	1	0.93 ¹	0.70 ¹	0.57	0.98 ¹	0.73	0.67
		3	0.45	0.46	0.34	0.49	0.35	0.30
ω_{11}	15	1	0.92 ¹	0.69 ¹	0.56	0.98 ¹	0.72	0.66
		3	0.45	0.46	0.33	0.49	0.35	0.30
ω_{12}	15	1	0.92 ¹	0.68 ¹	0.55	0.97 ¹	0.71	0.65
		3	0.45	0.46	0.33	0.49	0.35	0.30
ω_{13}	15	1	0.91 ¹	0.67 ¹	0.54	0.96 ¹	0.71	0.65
		3	0.45	0.46	0.33	0.50	0.35	0.30
ω_{14}	15	1	0.91 ¹	0.67 ¹	0.53	0.96 ¹	0.70	0.64
		3	0.45	0.46	0.33	0.50	0.35	0.31

Estimator	p	S	Prior 1			Prior 2		
			n			n		
			5	10	20	5	10	20
ω_1'	5	1	1.04 ¹ ₁	0.79 ¹ ₁	0.79 ¹ ₁	0.95 ¹ ₁	0.91 ¹ ₁	0.77 ¹ ₁
		3	0.95 ¹ ₁	0.69 ¹ ₁	0.64 ¹ ₁	0.72 ¹ ₁	0.57 ¹ ₁	0.50 ¹ ₁
		10	1.07 ¹ ₁	0.96 ¹ ₁	0.88 ¹ ₁	1.19 ¹ ₁	1.13 ¹ ₁	0.95 ¹ ₁
	15	1	0.76 ¹ ₂	0.82 ¹ ₁	0.56 ¹ ₁	0.78 ¹ ₁	0.61 ¹ ₁	0.54 ¹ ₁
		3	1.21 ¹ ₁	1.00 ¹ ₁	0.98 ¹ ₁	1.02 ¹ ₁	1.01 ¹ ₁	0.99 ¹ ₁
		3	0.60 ¹ ₁	0.59 ¹ ₁	0.52 ¹ ₁	0.74 ¹ ₁	0.79 ¹ ₁	0.48 ¹ ₁
ω_2'	5	1	0.92 ¹ ₁	0.65 ¹ ₁	0.71 ¹ ₁	0.86 ¹ ₁	0.82 ¹ ₁	0.66 ¹ ₁
		3	0.94 ¹ ₁	0.67 ¹ ₁	0.63 ¹ ₁	0.71 ¹ ₁	0.57 ¹ ₁	0.48 ¹ ₁
		10	0.98 ¹ ₁	0.87 ¹ ₁	0.78 ¹ ₁	1.08 ¹ ₁	1.02 ¹ ₁	0.84 ¹ ₁
	15	1	0.74 ¹ ₂	0.82 ¹ ₁	0.54 ¹ ₁	0.77 ¹ ₁	0.59 ¹ ₁	0.53 ¹ ₁
		3	1.13 ¹ ₁	0.92 ¹ ₁	0.89 ¹ ₁	0.95 ¹ ₁	0.92 ¹ ₁	0.91 ¹ ₁
		3	0.59 ¹ ₁	0.57 ¹ ₁	0.51 ¹ ₁	0.73 ¹ ₁	0.78 ¹ ₁	0.47 ¹ ₁
ω_3'	5	1	0.88 ¹ ₁	0.59 ¹ ₁	0.67 ¹ ₁	0.82 ¹ ₁	0.79 ¹ ₁	0.60 ¹ ₁
		3	0.94 ¹ ₁	0.67 ¹ ₁	0.62 ¹ ₁	0.71 ¹ ₁	0.59 ¹ ₁	0.47 ¹ ₁
		10	0.93 ¹ ₁	0.81 ¹ ₁	0.71 ¹ ₁	1.02 ¹ ₁	0.95 ¹ ₁	0.76 ¹ ₁
	15	1	0.72 ¹ ₂	0.82 ¹ ₁	0.53 ¹ ₁	0.76 ¹ ₁	0.58 ¹ ₁	0.52 ¹ ₁
		3	1.08 ¹ ₂	0.86 ¹ ₁	0.82 ¹ ₁	0.89 ¹ ₁	0.86 ¹ ₁	0.85 ¹ ₁
		3	0.58 ¹ ₁	0.56 ¹ ₁	0.50 ¹ ₁	0.72 ¹ ₁	0.77 ¹ ₁	0.46 ¹ ₁
ω_4'	5	1	0.85 ¹ ₁	0.55 ¹ ₁	0.64 ¹ ₁	0.80 ¹ ₁	0.77 ¹ ₁	0.56 ¹ ₁
		3	0.94 ¹ ₁	0.67 ¹ ₁	0.63 ¹ ₁	0.71 ¹ ₁	0.60 ¹ ₁	0.48 ¹ ₁
		10	0.89 ¹ ₁	0.77 ¹ ₁	0.66 ¹ ₁	0.97 ¹ ₁	0.90 ¹ ₁	0.71 ¹ ₁
	15	1	0.71 ¹ ₂	0.82 ¹ ₁	0.52 ¹ ₁	0.75 ¹ ₁	0.57 ¹ ₁	0.51 ¹ ₁
		3	1.04 ¹ ₁	0.81 ¹ ₁	0.77 ¹ ₁	0.85 ¹ ₁	0.82 ¹ ₁	0.80 ¹ ₁
		3	0.57 ¹ ₁	0.56 ¹ ₁	0.49 ¹ ₁	0.72 ¹ ₁	0.77 ¹ ₁	0.46 ¹ ₁
ω_5'	10	1	0.87 ¹ ₁	0.74 ¹ ₁	0.63 ¹ ₁	0.94 ¹ ₁	0.86 ¹ ₁	0.67 ¹ ₁
		3	0.71 ¹ ₂	0.82 ¹ ₁	0.51 ¹ ₁	0.75 ¹ ₁	0.57 ¹ ₁	0.51 ¹ ₁
	15	1	1.01 ¹ ₂	0.78 ¹ ₁	0.73 ¹ ₁	0.82 ¹ ₁	0.78 ¹ ₁	0.76 ¹ ₁
		3	0.57 ¹ ₁	0.55 ¹ ₁	0.48 ¹ ₁	0.71 ¹ ₁	0.76 ¹ ₁	0.45 ¹ ₁
ω_6'	10	1	0.85 ¹ ₁	0.72 ¹ ₁	0.60 ¹ ₁	0.91 ¹ ₁	0.83 ¹ ₁	0.63 ¹ ₁
		3	0.70 ¹ ₂	0.82 ¹ ₁	0.51 ¹ ₁	0.75 ¹ ₁	0.57 ¹ ₁	0.51 ¹ ₁
	15	1	0.99 ¹ ₂	0.75 ¹ ₁	0.69 ¹ ₁	0.79 ¹ ₁	0.75 ¹ ₁	0.73 ¹ ₁
		3	0.56 ¹ ₁	0.54 ¹ ₁	0.47 ¹ ₁	0.71 ¹ ₁	0.76 ¹ ₁	0.45 ¹ ₁
ω_7'	10	1	0.84 ¹ ₁	0.70 ¹ ₁	0.59 ¹ ₁	0.89 ¹ ₁	0.81 ¹ ₁	0.61 ¹ ₁
		3	0.70 ¹ ₁	0.82 ¹ ₁	0.50 ¹ ₁	0.75 ¹ ₁	0.57 ¹ ₁	0.51 ¹ ₁
	15	1	0.97 ¹ ₂	0.73 ¹ ₁	0.67 ¹ ₁	0.77 ¹ ₁	0.73 ¹ ₁	0.71 ¹ ₁
		3	0.56 ¹ ₁	0.54 ¹ ₁	0.47 ¹ ₁	0.71 ¹ ₁	0.76 ¹ ₁	0.44 ¹ ₁
ω_8'	10	1	0.83 ¹ ₁	0.68 ¹ ₁	0.58 ¹ ₁	0.88 ¹ ₁	0.79 ¹ ₁	0.59 ¹ ₁
		3	0.70 ¹ ₂	0.83 ¹ ₁	0.50 ¹ ₁	0.75 ¹ ₁	0.57 ¹ ₁	0.51 ¹ ₁
	15	1	0.96 ¹ ₂	0.71 ¹ ₁	0.64 ¹ ₁	0.75 ¹ ₁	0.71 ¹ ₁	0.69 ¹ ₁
		3	0.56 ¹ ₁	0.54 ¹ ₁	0.47 ¹ ₁	0.71 ¹ ₁	0.76 ¹ ₁	0.44 ¹ ₁
ω_9'	10	1	0.82 ¹ ₁	0.67 ¹ ₁	0.57 ¹ ₁	0.87 ¹ ₁	0.77 ¹ ₁	0.58 ¹ ₁
		3	0.70 ¹ ₂	0.83 ¹ ₁	0.50 ¹ ₁	0.75 ¹ ₁	0.57 ¹ ₁	0.52 ¹ ₁
	15	1	0.95 ¹ ₂	0.69 ¹ ₁	0.63 ¹ ₁	0.74 ¹ ₁	0.70 ¹ ₁	0.68 ¹ ₁
		3	0.56 ¹ ₁	0.54 ¹ ₁	0.46 ¹ ₁	0.71 ¹ ₁	0.76 ¹ ₁	0.44 ¹ ₁
ω_{10}'	15	1	0.94 ¹ ₂	0.68 ¹ ₁	0.61 ¹ ₁	0.73 ¹ ₁	0.69 ¹ ₁	0.67 ¹ ₁
		3	0.56 ¹ ₁	0.53 ¹ ₁	0.46 ¹ ₁	0.71 ¹ ₁	0.76 ¹ ₁	0.44 ¹ ₁

ω_r' , $r = 1, \dots, p-1$, are defined in (4.2.5(3)), b_r is set equal to 1 when $b_r < 0$.

Estimator	p	S	Prior 1			Prior 2		
			n			n		
			5	10	20	5	10	20
ω'_{11}	15	1	0.93 ²	0.67 ¹	0.60	0.72 ¹	0.68 ¹	0.66
		3	0.56	0.53	0.46	0.71 ¹	0.76 ¹	0.44
ω'_{12}	15	1	0.93 ²	0.66 ¹	0.59	0.71 ¹	0.67 ¹	0.65
		3	0.56	0.53	0.46	0.72 ¹	0.76 ¹	0.44
ω'_{13}	15	1	0.92 ²	0.66 ¹	0.58	0.70 ¹	0.66 ¹	0.65
		3	0.56	0.53	0.46	0.72 ¹	0.76 ¹	0.44
ω'_{14}	15	1	0.92 ²	0.65 ¹	0.57	0.70 ¹	0.66 ¹	0.64
		3	0.56	0.53	0.45	0.72 ¹	0.76 ¹	0.44
γ , mean	5	1	1.24 ²	0.99 ²	0.77 ¹	1.16 ¹	1.01 ¹	0.59
		3	0.86 ³	0.58 ²	0.45 ¹	0.66 ¹	0.65 ¹	0.35
	10	1	1.61 ¹	0.98 ¹	0.88	1.34 ¹	0.98 ¹	0.69
		3	0.61 ¹	0.61 ¹	0.39	0.87 ¹	0.55	0.40
γ , median	5	1	1.28 ²	1.01 ²	0.77 ²	1.16 ¹	1.02 ¹	0.60
		3	0.86 ³	0.55 ²	0.44 ¹	0.66 ¹	0.63 ¹	0.34
	10	1	1.62 ¹	0.98 ¹	0.84 ¹	1.33 ¹	0.97 ¹	0.68
		3	0.64 ¹	0.59 ¹	0.36	0.85 ¹	0.54	0.40
γ , s, m	5	1	1.03 ²	0.76 ¹	0.70 ¹	1.01 ¹	0.92 ¹	0.58
		3	0.73 ¹	0.54 ¹	0.45 ¹	0.63 ¹	0.70 ¹	0.36
	10	1	1.33 ²	0.77 ¹	0.74 ¹	1.01 ¹	0.81 ¹	0.61
		3	0.52	0.56 ¹	0.37	0.72 ¹	0.45	0.42
γ , s, md	5	1	1.02 ²	0.75 ¹	0.74 ¹	0.97 ¹	1.09 ²	0.63
		3	0.69 ¹	0.58 ¹	0.50 ²	0.62 ¹	0.76 ¹	0.39 ¹
	10	1	1.36 ²	0.88 ²	0.94 ¹	1.08 ¹	0.96 ¹	0.67 ¹
		3	0.56	0.84 ¹	0.56 ¹	0.76 ¹	0.62 ¹	0.50 ¹
$\gamma^{(c)}$, mean	5	1	0.66 ¹	0.62 ¹	0.56 ¹	0.79 ¹	0.83 ¹	0.55
		3	0.66 ¹	0.46	0.39	0.55	0.54 ¹	0.33
	10	1	0.74 ¹	0.62 ¹	0.64 ¹	0.77 ¹	0.79 ¹	0.60
		3	0.54 ¹	0.49 ¹	0.33	0.66	0.44	0.38
$\gamma^{(c)}$, median	5	1	0.69 ¹	0.62 ¹	0.56 ¹	0.80 ¹	0.85 ¹	0.55
		3	0.68 ¹	0.43	0.38 ¹	0.55 ¹	0.52 ¹	0.33
	10	1	0.76 ¹	0.63 ¹	0.61	0.83	0.78 ¹	0.60
		3	0.55 ¹	0.49 ¹	0.31	0.63	0.44	0.39
$\gamma^{(c)}$, s, m	5	1	0.78 ¹	0.68 ¹	0.59 ¹	1.05 ²	0.98 ¹	0.66 ¹
		3	0.77 ¹	0.63 ¹	0.47 ¹	0.73 ²	0.78 ¹	0.49 ¹
	10	1	1.00 ¹	0.77 ¹	0.89 ²	1.13 ²	0.95 ¹	0.63 ¹
		3	0.73 ¹	0.74 ²	0.46 ¹	1.05 ²	0.65	0.59 ¹
$\gamma^{(c)}$, s, md	5	1	0.82 ¹	0.69 ¹	0.63 ¹	1.09 ²	1.16 ²	0.71 ¹
		3	0.78 ²	0.68 ¹	0.51 ¹	0.76 ¹	0.84 ¹	0.52 ¹
	10	1	1.17 ¹	0.97 ³	1.09 ²	1.28 ²	1.16 ²	0.71 ¹
		3	0.84 ¹	1.10 ³	0.66 ¹	1.24 ²	0.90 ¹	0.68 ¹

- , mean denotes the estimator which is constructed by the procedure defined in (5.2.1 point 1.).
- , median denotes the estimator which is constructed by the procedure defined in (5.2.1 point 2.).
- , s, m denotes the estimator which is constructed by the procedure defined in (5.2.1 point 3.).
- , s, md denotes the estimator which is constructed by the procedure defined in (5.2.1 point 4.).

Estimator	p	S	Prior 1			Prior 2		
			n			n		
			5	10	20	5	10	20
ϵ , mean	5	1	1.34 ²	0.94 ²	0.75 ¹	1.44 ¹	0.99 ¹	0.59
		3	1.15 ²	0.60 ¹	0.45 ¹	0.86 ¹	0.68 ¹	0.36
	10	1	2.11 ³	1.14 ²	0.81 ¹	1.50 ²	1.10 ¹	0.69
		3	1.20 ²	0.80 ¹	0.42	1.28 ¹	0.62	0.43
ϵ , median	5	1	1.31 ²	0.92 ²	0.74 ¹	1.12 ¹	0.97 ¹	0.59
		3	0.98 ¹	0.55 ¹	0.43 ¹	0.75 ¹	0.64 ¹	0.34
	10	1	1.57 ³	0.97 ¹	0.73 ¹	1.29 ²	0.95 ¹	0.68
		3	0.86 ¹	0.63 ¹	0.35	0.93 ¹	0.50	0.40
ϵ , s, m	5	1	1.18 ²	0.86 ¹	0.67 ¹	1.21 ²	0.98 ¹	0.58
		3	0.94 ³	0.55 ²	0.46 ²	0.75 ³	0.73 ²	0.34 ¹
	10	1	2.40 ²	0.99 ²	0.93 ²	1.48 ²	1.05 ²	0.66 ¹
		3	0.87 ²	0.84 ²	0.44	1.00 ²	0.69 ¹	0.45
ϵ , s, md	5	1	1.40 ²	0.90 ¹	0.72 ¹	1.48 ³	1.17 ²	0.69 ¹
		3	1.13 ²	0.63 ¹	0.49 ³	1.00 ²	0.78 ¹	0.37 ¹
	10	1	1.45 ³	1.50 ³	1.20 ³	2.31 ³	1.71 ³	0.76 ¹
		3	1.96 ³	1.66 ³	0.67 ¹	2.24 ³	1.27 ³	0.61 ¹
ϕ , mean	5	1	0.85 ¹	0.62 ¹	0.61 ¹	0.87 ¹	0.84 ¹	0.59
		3	0.72 ¹	0.44 ¹	0.39	0.62 ¹	0.49	0.30
	10	1	0.91 ¹	0.68 ¹	0.62 ¹	0.81 ¹	0.83	0.64
		3	0.57 ¹	0.48	0.29	0.60	0.39	0.36
ϕ , median	5	1	0.96 ²	0.62 ¹	0.61 ¹	0.94 ¹	0.83	0.58
		3	0.82 ¹	0.45 ¹	0.39	0.69 ¹	0.51 ¹	0.31
	10	1	1.04 ¹	0.72 ¹	0.61 ¹	0.91 ¹	0.72 ¹	0.63
		3	0.57 ¹	0.51	0.30	0.65	0.46 ¹	0.37
ϕ , s, m	5	1	0.99 ²	0.55 ¹	0.56 ¹	1.42 ³	1.05 ³	0.52
		3	0.88 ²	0.50 ¹	0.40	0.84 ¹	0.52	0.32
	10	1	0.86 ¹	0.64 ¹	0.58 ¹	0.90 ¹	0.73 ¹	0.56
		3	0.54 ¹	0.46	0.32	0.75 ¹	0.46 ¹	0.36
ϕ , s, md	5	1	1.00 ²	0.54 ¹	0.56 ¹	1.42 ³	1.04 ³	0.53
		3	0.91 ¹	0.50 ¹	0.40	0.85 ¹	0.54 ¹	0.32
	10	1	0.86 ¹	0.66 ¹	0.57 ¹	0.92 ¹	0.72 ¹	0.53
		3	0.56 ¹	0.47	0.33	0.75 ¹	0.46 ¹	0.35
ϕ' , mean	5	1	0.82 ¹	0.62 ¹	0.61	0.80 ¹	0.83 ¹	0.59
		3	0.68 ¹	0.42 ¹	0.39 ¹	0.57 ¹	0.46 ¹	0.30
	10	1	0.89 ¹	0.68 ¹	0.62 ¹	0.80 ¹	0.82 ¹	0.64
		3	0.55 ¹	0.48	0.29	0.58	0.39	0.36
ϕ' , median	5	1	0.89 ¹	0.62 ¹	0.61 ¹	0.84 ¹	0.83 ¹	0.58
		3	0.76 ²	0.42 ¹	0.39 ¹	0.60 ¹	0.48 ¹	0.31
	10	1	1.03 ¹	0.71 ¹	0.61 ¹	0.89 ¹	0.82 ¹	0.63
		3	0.55 ¹	0.51	0.30	0.62	0.46 ¹	0.37
ϕ' , s, m	5	1	0.96 ²	0.56 ¹	0.56 ¹	1.35 ³	1.03 ³	0.52
		3	0.84 ¹	0.48 ¹	0.40 ¹	0.78 ¹	0.50 ¹	0.32
	10	1	0.85 ¹	0.64 ¹	0.58 ¹	0.91 ¹	0.74 ¹	0.56
		3	0.53 ¹	0.45	0.32	0.72 ¹	0.46 ¹	0.36

Estimator	p	S	Prior 1			Prior 2		
			n			n		
			5	10	20	5	10	20
ϕ', s, md	5	1	0.97 ² ₂	0.55 ¹	0.57 ¹	1.35 ³	1.01 ³	0.53
		3	0.87 ¹ ₁	0.49 ¹	0.40 ¹ ₁	0.80 ¹ ₁	0.52 ¹ ₁	0.32
	10	1	0.85 ¹ ₁	0.66 ¹	0.57 ¹ ₁	0.94 ¹ ₁	0.72 ¹ ₁	0.54
		3	0.52 ¹ ₁	0.46 ¹	0.32 ¹	0.73 ¹ ₁	0.46 ¹ ₁	0.36
$\phi'', mean$	5	1	0.69 ¹ ₁	0.60 ¹	0.61 ¹	0.79 ¹	0.85 ¹	0.59
		3	0.61 ¹ ₁	0.41 ¹	0.39 ¹	0.56 ¹	0.46 ¹ ₁	0.30
	10	1	0.82 ¹ ₁	0.67 ¹	0.61 ¹	0.77 ¹ ₁	0.81 ¹ ₁	0.64
		3	0.54 ¹ ₁	0.47 ¹	0.29 ¹	0.58 ¹	0.39 ¹	0.36
$\phi'', median$	5	1	0.73 ¹ ₁	0.61 ¹	0.61 ¹	0.76 ¹ ₁	0.84 ¹ ₁	0.58
		3	0.66 ¹ ₁	0.42 ¹	0.39 ¹	0.59 ¹ ₁	0.48 ¹ ₁	0.30
	10	1	0.92 ¹ ₁	0.70 ¹	0.61 ¹	0.85 ¹ ₁	0.81 ¹ ₁	0.63
		3	0.55 ¹ ₁	0.50 ¹	0.30 ¹	0.62 ¹	0.45 ¹ ₁	0.37
ϕ'', s, m	5	1	0.81 ¹ ₂	0.54 ¹	0.56 ¹	1.25 ³ ₁	0.99 ³ ₁	0.52
		3	0.77 ¹ ₁	0.49 ¹ ₁	0.39 ¹	0.77 ¹ ₁	0.50 ¹ ₁	0.32
	10	1	0.79 ¹ ₁	0.62 ¹	0.57 ¹ ₁	0.90 ¹ ₁	0.73 ¹ ₁	0.56
		3	0.54 ¹ ₁	0.44 ¹	0.32 ¹	0.72 ¹ ₁	0.45 ¹ ₁	0.35
ϕ'', s, md	5	1	0.83 ¹ ₂	0.53 ¹	0.56 ¹	1.26 ³ ₁	0.98 ³ ₁	0.53
		3	0.79 ¹ ₁	0.49 ¹	0.39 ¹	0.78 ¹ ₁	0.52 ¹ ₁	0.31
	10	1	0.78 ¹ ₁	0.63 ¹	0.57 ¹ ₁	0.93 ¹ ₁	0.71 ¹ ₁	0.53
		3	0.54 ¹ ₁	0.45 ¹	0.32 ¹	0.72 ¹ ₁	0.45 ¹ ₁	0.35
$\psi, mean$	5	1	0.84 ¹ ₁	0.62 ¹	0.62 ¹	0.85 ¹ ₁	0.86 ¹ ₁	0.58
		3	0.72 ¹ ₁	0.42 ¹ ₁	0.39 ¹	0.62 ¹ ₁	0.48 ¹ ₁	0.30
	10	1	0.91 ¹ ₁	0.68 ¹	0.62 ¹	0.81 ¹ ₁	0.82 ¹ ₁	0.64
		3	0.56 ¹ ₁	0.48 ¹	0.29 ¹	0.60 ¹	0.39 ¹	0.36
$\psi, median$	5	1	0.93 ¹ ₁	0.64 ¹	0.61 ¹	0.91 ¹ ₁	0.86 ¹ ₁	0.57
		3	0.80 ¹ ₂	0.42 ¹ ₁	0.39 ¹	0.68 ¹ ₁	0.51 ¹ ₁	0.31
	10	1	1.01 ¹ ₂	0.71 ¹	0.61 ¹	0.88 ¹ ₁	0.81 ¹ ₁	0.62
		3	0.56 ¹ ₁	0.51 ¹	0.30 ¹	0.64 ¹	0.44 ¹ ₁	0.37
ψ, s, m	5	1	0.95 ¹ ₂	0.57 ¹	0.58 ¹	1.36 ³ ₁	1.07 ³ ₁	0.51
		3	0.88 ¹ ₁	0.47 ¹ ₁	0.40 ¹	0.83 ¹ ₁	0.51 ¹ ₁	0.32
	10	1	0.84 ¹ ₁	0.64 ¹	0.58 ¹	0.89 ¹ ₁	0.73 ¹ ₁	0.55
		3	0.54 ¹ ₁	0.45 ¹	0.32 ¹	0.72 ¹ ₁	0.45 ¹ ₁	0.36
ψ, s, md	5	1	0.96 ¹ ₂	0.57 ¹	0.58 ¹	1.34 ³ ₁	1.06 ³ ₁	0.52
		3	0.90 ¹ ₁	0.47 ¹	0.39 ¹	0.84 ¹ ₁	0.52 ¹ ₁	0.32
	10	1	0.85 ¹ ₁	0.64 ¹	0.57 ¹ ₁	0.93 ¹ ₁	0.72 ¹ ₁	0.53
		3	0.54 ¹ ₁	0.45 ¹	0.32 ¹	0.71 ¹ ₁	0.45 ¹ ₁	0.36
$\omega, mean$	5	1	0.98 ¹ ₁	0.73 ¹	0.64 ¹	0.90 ¹	0.87 ¹	0.63
		3	0.61 ¹ ₂	0.40 ¹	0.38 ¹	0.54 ¹ ₁	0.45 ¹ ₁	0.30
	10	1	1.04 ¹ ₁	0.76 ¹	0.65 ¹	1.00 ¹ ₁	0.87 ¹ ₁	0.69
		3	0.53 ¹ ₁	0.45 ¹	0.51 ¹ ₁	0.54 ¹ ₁	0.36 ¹ ₁	0.36
	15	1	0.96 ¹ ₁	0.75 ¹	0.63 ¹ ₁	1.03 ¹ ₁	0.78 ¹ ₁	0.71
		3	0.46 ¹	0.47 ¹	0.34 ¹	0.49 ¹	0.35 ¹	0.31
$\omega, median$	5	1	0.97 ¹ ₁	0.72 ¹	0.64 ¹	0.89 ¹ ₁	0.87 ¹ ₁	0.62
		3	0.61 ¹ ₂	0.40 ¹	0.38 ¹	0.54 ¹ ₁	0.46 ¹ ₁	0.30
	10	1	1.04 ¹ ₁	0.75 ¹	0.64 ¹ ₁	0.99 ¹ ₁	0.86 ¹ ₁	0.67
		3	0.53 ¹ ₁	0.45 ¹	0.29 ¹	0.54 ¹ ₁	0.36 ¹ ₁	0.36
	15	1	0.96 ¹ ₁	0.73 ¹	0.62 ¹	1.02 ¹ ₁	0.77 ¹ ₁	0.70
		3	0.46 ¹	0.46 ¹	0.34 ¹	0.49 ¹	0.35 ¹	0.31

Estimator	p	S	Prior 1			Prior 2		
			n			n		
			5	10	20	5	10	20
ω, s, m	5	1	0.93 ¹	0.65 ¹	0.59 ¹	0.86 ¹	0.83 ¹	0.56
		3	0.61 ¹	0.40 ¹	0.38 ¹	0.54 ¹	0.48 ¹	0.30
	10	1	0.99 ¹	0.69 ¹	0.59 ¹	0.93 ¹	0.78 ¹	0.60
		3	0.52 ¹	0.45 ¹	0.28	0.54 ¹	0.36	0.37
	15	1	0.91 ¹	0.67 ¹	0.54	0.96 ¹	0.71	0.65
		3	0.45	0.46	0.33	0.50	0.35	0.31
ω, s, md	5	1	0.92 ¹	0.64 ¹	0.60 ¹	0.86 ¹	0.83 ¹	0.57
		3	0.61 ¹	0.40 ¹	0.38 ¹	0.54 ¹	0.48 ¹	0.30
	10	1	0.99 ²	0.69 ¹	0.59 ¹	0.93 ¹	0.77 ¹	0.58
		3	0.52 ¹	0.46 ¹	0.28	0.54 ¹	0.36	0.37
	15	1	0.91 ¹	0.67 ¹	0.53	0.95 ¹	0.70	0.64
		3	0.46	0.46	0.33	0.50	0.35	0.31
ω', mean	5	1	0.90 ¹	0.63 ¹	0.69 ¹	0.84 ¹	0.81 ¹	0.63 ¹
		3	0.94 ¹	0.67 ¹	0.62 ¹	0.71 ¹	0.58 ¹	0.48 ¹
	10	1	0.87 ¹	0.75 ¹	0.64 ¹	0.95 ¹	0.87 ¹	0.68 ¹
		3	0.71 ¹	0.82 ¹	0.51 ¹	0.75 ¹	0.57 ¹	0.51 ¹
	15	1	0.97 ²	0.73 ¹	0.67 ¹	0.77 ¹	0.73 ¹	0.71
		3	0.56	0.54	0.47	0.71 ¹	0.76 ¹	0.44
ω', median	5	1	0.90 ¹	0.62 ¹	0.69 ¹	0.84 ¹	0.80 ¹	0.62 ¹
		3	0.94 ¹	0.67 ¹	0.62 ¹	0.71 ¹	0.58 ¹	0.48 ¹
	10	1	0.87 ¹	0.74 ¹	0.63 ¹	0.94 ¹	0.86 ¹	0.67 ¹
		3	0.71 ¹	0.82 ¹	0.51 ¹	0.75 ¹	0.57 ¹	0.51 ¹
	15	1	0.97 ²	0.72 ¹	0.65 ¹	0.76 ¹	0.72 ¹	0.70
		3	0.56	0.54	0.47	0.71 ¹	0.76 ¹	0.44
ω', s, m	5	1	0.85 ¹	0.55 ¹	0.64 ¹	0.80 ¹	0.77 ¹	0.57 ¹
		3	0.94 ¹	0.67 ¹	0.63 ¹	0.71 ¹	0.60 ¹	0.48 ¹
	10	1	0.82 ¹	0.67 ¹	0.58 ¹	0.88 ¹	0.78 ¹	0.59 ¹
		3	0.70 ²	0.82 ¹	0.50 ¹	0.75 ¹	0.57 ¹	0.52 ¹
	15	1	0.92	0.66	0.58	0.70	0.66	0.65
		3	0.56	0.54	0.46	0.72 ¹	0.76 ¹	0.44
ω', s, md	5	1	0.85 ¹	0.55 ¹	0.64 ¹	0.80 ¹	0.76 ¹	0.58 ¹
		3	0.94 ¹	0.67 ¹	0.63 ¹	0.71 ¹	0.60 ¹	0.48 ¹
	10	1	0.82 ¹	0.68 ¹	0.57 ¹	0.87 ¹	0.77 ¹	0.58 ¹
		3	0.70 ²	0.83 ¹	0.50 ¹	0.75 ¹	0.57 ¹	0.52 ¹
	15	1	0.92	0.65	0.57	0.70 ¹	0.66	0.64
		3	0.56	0.53	0.46	0.72 ¹	0.76 ¹	0.44
Robbins	5	1	2.04 ³	2.05 ³	1.44 ³	2.08 ³	1.81 ²	1.23 ¹
		3	1.56 ³	1.27 ³	1.26 ³	1.60 ³	1.53 ³	0.90 ³
	10	1	3.38 ³	2.53 ³	3.30 ³	2.95 ³	2.46 ³	1.96 ³
		3	2.32 ³	2.43 ³	2.17 ³	2.61 ³	1.96 ³	1.42 ³
	15	1	4.38 ³	3.10 ³	3.42 ³	4.31 ³	4.20 ³	4.22 ³
		3	3.02 ³	3.27 ³	1.92 ³	2.74 ³	2.75 ³	2.50 ³
Robbins, P	5	1	2.34 ³	2.49 ³	1.68 ³	2.29 ³	2.91 ³	2.25 ³
		3	2.04 ³	2.15 ³	1.54 ³	2.27 ³	2.23 ³	1.33 ³
	10	1	3.80 ³	2.68 ³	4.96 ³	3.18 ³	3.34 ³	3.30 ³
		3	2.37 ³	3.00 ³	2.10 ³	2.96 ³	3.02 ³	1.85 ³
	15	1	4.70 ³	4.32 ³	5.77 ³	4.73 ³	5.88 ³	5.41 ³
		3	3.85 ³	3.22 ³	4.90 ³	2.94 ³	3.86 ³	3.16 ³

Robbins is defined in (5.3(2)).

Robbins, P is defined in (5.3(3)).

Estimator	p	S	Prior 1			Prior 2		
			n			n		
			5	10	20	5	10	20
Parametric	5	1	0.70 ¹	0.49 ¹	0.51 ¹	0.67 ¹	0.69 ¹	0.53
		3	0.72 ¹	0.62 ¹	0.57 ¹	0.60 ¹	0.58 ¹	0.40
	10	1	0.69 ¹	0.56 ¹	0.57 ¹	0.75 ¹	0.67 ¹	0.56
		3	0.60 ¹	0.64 ¹	0.38	0.62	0.48	0.44
	15	1	0.67 ¹	0.59 ¹	0.51	0.58	0.58	0.65
		3	0.47	0.52	0.38	0.50	0.44	0.44
Parametric, P	5	1	0.95 ¹	0.99 ¹	0.83 ¹	0.93 ¹	1.03	0.93
		3	0.49 ¹	0.41 ¹	0.40 ¹	0.44 ¹	0.40 ¹	0.33
	10	1	0.99 ¹	0.92 ¹	0.97 ¹	1.01 ¹	1.08 ¹	1.02
		3	0.48 ¹	0.39 ¹	0.34 ¹	0.45 ¹	0.37	0.37
	15	1	0.94 ¹	0.93 ¹	0.96 ¹	0.97 ¹	0.97	1.00
		3	0.39	0.45	0.37	0.39	0.33	0.33
Ergle	5	1	1.77 ³	1.20 ¹	1.25 ¹	1.99 ³	1.70 ²	1.74 ¹
		3	0.87 ¹	0.79 ¹	0.69	0.79 ¹	0.74 ¹	0.65
	10	1	0.94 ¹	0.97 ¹	0.86 ¹	1.17 ¹	1.09 ¹	0.85 ¹
		3	0.73 ¹	0.74 ¹	0.60	0.78 ¹	0.65 ¹	0.64 ¹
	15	1	1.01 ¹	0.88 ¹	0.86 ¹	0.93 ¹	0.94 ¹	0.90 ¹
		3	0.62	0.64 ¹	0.70	0.79 ¹	0.70 ¹	0.69 ¹
Lemon	5	1	1.49 ³	0.87 ¹	0.86 ¹	1.90 ³	1.33 ¹	1.19
		3	0.84 ¹	0.65 ¹	0.57 ¹	0.72 ¹	0.68 ¹	0.60 ¹
	10	1	0.75 ¹	0.59 ¹	0.59	0.88 ¹	0.70	0.58 ¹
		3	0.54 ¹	0.57 ¹	0.38	0.65 ¹	0.50 ¹	0.47 ¹
	15	1	0.85 ¹	0.66 ¹	0.53	0.73 ¹	0.62 ¹	0.72 ¹
		3	0.46	0.50 ¹	0.40	0.61 ¹	0.53 ¹	0.44
Lemon, it.	5	1	1.50 ³	0.96 ¹	0.94 ¹	1.86 ³	1.36 ¹	1.30 ¹
		3	0.76 ¹	0.57	0.49	0.64 ¹	0.55	0.49 ¹
	10	1	0.71 ¹	0.64 ¹	0.59	0.84 ¹	0.78 ¹	0.65 ¹
		3	0.51 ¹	0.50 ¹	0.29	0.55 ¹	0.42 ¹	0.39 ¹
	15	1	0.82 ¹	0.62 ¹	0.51	0.77 ¹	0.59	0.66
		3	0.38	0.43 ¹	0.29	0.51	0.38	0.31
Lemon, P	5	1	1.63 ³	0.98 ¹	1.01 ¹	1.94 ³	1.50 ¹	1.41 ¹
		3	0.85 ¹	0.67	0.60	0.73 ¹	0.67 ¹	0.59
	10	1	0.72 ¹	0.62 ¹	0.58	0.87 ¹	0.74	0.59
		3	0.53 ¹	0.52 ¹	0.35	0.59	0.45	0.42 ¹
	15	1	0.79 ¹	0.61 ¹	0.51	0.69 ¹	0.59	0.66 ¹
		3	0.44	0.45 ¹	0.36	0.55 ¹	0.45	0.39
Lemon, it., P	5	1	1.71 ³	1.17 ¹	1.20 ¹	1.94 ³	1.62 ²	1.62 ¹
		3	0.82 ¹	0.66	0.57	0.69 ¹	0.59	0.53
	10	1	0.72 ¹	0.72 ¹	0.65	0.87 ¹	0.90	0.74
		3	0.50 ¹	0.46 ¹	0.30	0.50	0.38	0.37
	15	1	0.77 ¹	0.61	0.53	0.75 ¹	0.60	0.64
		3	0.38	0.39	0.29	0.45	0.32	0.29

Parametric is defined in (5.3(4)).

Parametric, P is defined in (5.3(5)).

Ergle is defined in (5.3(10)).

Lemon is defined in (5.3(12)).

Lemon, it. is defined in (5.3(13)).

Lemon, P is defined in (5.3(14)).

Lemon, it., P is defined in (5.3(15)).

Estimator	p	S	Prior 1			Prior 2		
			n			n		
			5	10	20	5	10	20
Griffin a	5	1	1.67 ³ ₁	1.00 ¹ ₁	1.02 ¹ ₁	2.03 ³ ₁	1.52 ¹ ₁	1.43 ¹ ₁
		3	0.97 ¹ ₁	0.83 ¹ ₁	0.73 ¹ ₁	0.83 ¹ ₁	0.86 ¹ ₁	0.78 ¹ ₁
	10	1	0.84 ¹ ₁	0.67 ¹ ₁	0.68 ¹ ₁	0.97 ¹ ₁	0.77 ¹ ₁	0.62 ¹ ₁
		3	0.60 ¹ ₁	0.64 ¹ ₁	0.45 ¹ ₁	0.72 ¹ ₁	0.56 ¹ ₁	0.54 ¹ ₁
	15	1	0.88 ¹ ₁	0.70 ¹ ₁	0.58 ¹ ₁	0.76 ¹ ₁	0.67 ¹ ₁	0.76 ¹ ₁
		3	0.49 ¹ ₁	0.53 ¹ ₁	0.45 ¹ ₁	0.66 ¹ ₁	0.57 ¹ ₁	0.48 ¹ ₁
Griffin b	5	1	0.72 ¹ ₁	0.50 ¹ ₁	0.51 ¹ ₁	0.70 ¹ ₁	0.70 ¹ ₁	0.53 ¹ ₁
		3	0.78 ¹ ₁	0.64 ¹ ₁	0.58 ¹ ₁	0.64 ¹ ₁	0.59 ¹ ₁	0.40 ¹ ₁
	10	1	0.69 ¹ ₁	0.56 ¹ ₁	0.57 ¹ ₁	0.76 ¹ ₁	0.68 ¹ ₁	0.56 ¹ ₁
		3	0.62 ¹ ₁	0.65 ¹ ₁	0.39 ¹ ₁	0.62 ¹ ₁	0.49 ¹ ₁	0.44 ¹ ₁
	15	1	0.68 ¹ ₁	0.60 ¹ ₁	0.51 ¹ ₁	0.59 ¹ ₁	0.58 ¹ ₁	0.65 ¹ ₁
		3	0.49 ¹ ₁	0.52 ¹ ₁	0.39 ¹ ₁	0.51 ¹ ₁	0.44 ¹ ₁	0.44 ¹ ₁
Griffin b, P	5	1	0.80 ¹ ₁	0.55 ¹ ₁	0.63 ¹ ₁	0.69 ¹ ₁	0.73 ¹ ₁	0.68 ¹ ₁
		3	0.95 ¹ ₁	0.71 ¹ ₁	0.78 ¹ ₁	0.66 ¹ ₁	0.69 ¹ ₁	0.63 ¹ ₁
	10	1	0.78 ¹ ₁	0.62 ¹ ₁	0.66 ¹ ₁	0.74 ¹ ₁	0.82 ¹ ₁	0.66 ¹ ₁
		3	0.70 ¹ ₁	0.72 ¹ ₁	0.71 ¹ ₁	0.66 ¹ ₁	0.54 ¹ ₁	0.60 ¹ ₁
	15	1	0.75 ¹ ₁	0.61 ¹ ₁	0.62 ¹ ₁	0.65 ¹ ₁	0.63 ¹ ₁	0.73 ¹ ₁
		3	0.59 ¹ ₁	0.61 ¹ ₁	0.56 ¹ ₁	0.68 ¹ ₁	0.64 ¹ ₁	0.56 ¹ ₁
Martz	5	1	0.59 ¹ ₁	0.47 ¹ ₁	0.53 ¹ ₁	0.68 ¹ ₁	0.66 ¹ ₁	0.58 ¹ ₁
		3	0.51 ¹ ₁	0.47 ¹ ₁	0.47 ¹ ₁	0.47 ¹ ₁	0.49 ¹ ₁	0.47 ¹ ₁
	10	1	0.58 ¹ ₁	0.52 ¹ ₁	0.61 ¹ ₁	0.69 ¹ ₁	0.59 ¹ ₁	0.47 ¹ ₁
		3	0.47 ¹ ₁	0.50 ¹ ₁	0.42 ¹ ₁	0.55 ¹ ₁	0.40 ¹ ₁	0.46 ¹ ₁
	15	1	0.70 ¹ ₁	0.60 ¹ ₁	0.51 ¹ ₁	0.59 ¹ ₁	0.60 ¹ ₁	0.65 ¹ ₁
		3	0.41 ¹ ₁	0.44 ¹ ₁	0.45 ¹ ₁	0.49 ¹ ₁	0.47 ¹ ₁	0.45 ¹ ₁
Martz, P	5	1	0.63 ¹ ₁	0.52 ¹ ₁	0.62 ¹ ₁	0.67 ¹ ₁	0.70 ¹ ₁	0.65 ¹ ₁
		3	0.48 ¹ ₁	0.45 ¹ ₁	0.47 ¹ ₁	0.46 ¹ ₁	0.46 ¹ ₁	0.43 ¹ ₁
	10	1	0.58 ¹ ₁	0.56 ¹ ₁	0.62 ¹ ₁	0.68 ¹ ₁	0.62 ¹ ₁	0.48 ¹ ₁
		3	0.47 ¹ ₁	0.47 ¹ ₁	0.40 ¹ ₁	0.52 ¹ ₁	0.39 ¹ ₁	0.42 ¹ ₁
	15	1	0.69 ¹ ₁	0.58 ¹ ₁	0.51 ¹ ₁	0.58 ¹ ₁	0.60 ¹ ₁	0.61 ¹ ₁
		3	0.40 ¹ ₁	0.41 ¹ ₁	0.42 ¹ ₁	0.46 ¹ ₁	0.43 ¹ ₁	0.42 ¹ ₁
Estimated average squared error of the ML estimator	5	1	.0264	.0259	.0272	.0257	.0276	.0264
		3	.0288	.0298	.0295	.0298	.0307	.0307
	10	1	.0148	.0158	.0143	.0142	.0139	.0142
		3	.0165	.0157	.0153	.0151	.0156	.0154
	15	1	.0097	.0104	.0099	.0094	.0098	.0098
		3	.0109	.0102	.0106	.0111	.0106	.0106

Griffin a is defined in (5.3(16)).

Griffin b is defined in (5.3(17)).

Griffin b, P is defined in (5.3(30)).

Martz is defined in (5.3(31)).

Martz, P is defined in (5.3(35)).

Estimated average squared error of the ML estimator is calculated from the denominator of (5.1(2)).

REFERENCES

- Bennett G.K. & H.F. Martz (1972), A continuous empirical Bayes smoothing technique, *Biometrika* 59, p. 361-68.
- Ergle W.D. & R.G. Krutchkoff (1972), A non-parametric empirical Bayes estimator, *Biometrika* 59, p. 483-85.
- Gabbert J.T. & R.G. Krutchkoff (1970), Supplementary sample non-parametric empirical Bayes in a quality control situation, *Biometrika* 57, p. 67-69.
- Griffin B.S. & R.G. Krutchkoff (1971 a), An empirical Bayes estimator for $P(\text{success})$ in the binomial distribution, *Sankyā Ser. B* 33, p. 217-24. (1971 b), Optimal linear estimators: an empirical Bayes version with application to the binomial distribution, *Biometrika* 58, p. 195-201.
- Johns M.V. Jr (1957), Non-parametric Bayes Procedures, *Ann. Math. Statist.* 28, p. 649-69.
- Krutchkoff R.G. (1967), A supplementary sample non-parametric empirical Bayes approach to some statistical decision problems, *Biometrika* 54, p. 451-58.
- Lehmann E.L. (1959), *Testing Statistical Hypotheses*, Wiley, New York.
- Lemon G.H. & R.G. Krutchkoff (1969), An empirical Bayes smoothing technique, *Biometrika* 56, p. 361-65.
- Lindley D.V. (1965), *Introduction to Probability and Statistics*, Part 2, Cambridge Univ. Press.
- Lundqvist E. & A. Sjöberg (1969), *Numerisk analys 1*, Studentlitt. Lund.
- Maritz J.S. (1966), Smooth empirical Bayes estimation for one-parameter discrete distributions, *Biometrika* 53, p. 417-429.
- (1967), Smooth empirical Bayes estimation for continuous distribution, *Biometrika* 54, p. 435-50.
- (1969), Empirical Bayes estimation for the Poisson distribution, *Biometrika* 56, p. 349-59.
- (1970), *Empirical Bayes Methods*, Methuen, London.
- Parzen E. (1962), On estimation of a probability density function and mode, *Ann. Math. Statist.* 33, p. 1065-76.
- Raiffa H. & R. Schlaifer (1961), *Applied statistical decision theory*, Harvard University, Boston.
- Robbins H. (1955), An empirical Bayes approach to statistics, *Proc. Second Berkeley Symp. Math. Statist. Prob.*, 1, p. 157-164. University of California Press.
- (1964), The empirical Bayes approach to statistical decision problems, *Ann. Math. Statist.* 35, p. 1-20.
- Rutherford J.R. & R.G. Krutchkoff (1969), Some empirical Bayes techniques in point estimation, *Biometrika* 56, p. 133-37.

INDEX

See also the table of contents.

- asymptotic optimality 15
- average estimator 28
- Bayes
 - decision function 11
 - estimator 12
 - methods 11
 - risk 13
- Bennett and Martz (1972) 26, 109, 112
- Bernoulli case 74
- choise of r 68, 101
- Clemmer and Krutchkoff (1968) 26
- conditional distribution 14
- combinatorial argument 83
- current data 14
- data
 - current 14
 - past 13
 - matrix 40, 99
- density estimator 26, 32, 105, 109
- discrete conditional distribution 41
- Empirical Bayes (EB)
 - estimators 17, 102
 - methods 13
 - non-parametric estimators 27, 39, 75, 101
 - selection of estimators 14
- Ergle and Krutchkoff (1972) 32, 104, 112, 113
- 'Ergle' 128
- extra risk 15
- function u 42
- function v 63
- Gabbert and Krutchkoff (1970) 32
- Griffin and Krutchkoff (1971 a) 22, 107, 112
 - (1971 b) 24, 38, 107, 112
- 'Griffin a' 129
- 'Griffin b' 129
- 'Griffin b, P' 129
- identifiability 14
- inaccuracy 44
- indicators 49, 50
- iteration 37, 99
- Johns (1957) 27, 35, 41, 46, 48, 50, 51, 111, 114
- Krutchkoff (1967) 30, 35, 62, 111

- least squares estimators 63
- Lehmann (1959) 13
- Lemon and Krutchkoff (1969) 21, 22, 26, 32, 38, 105, 112
- 'Lemon' 128
- 'Lemon, it.' 128
- 'Lemon, P' 128
- 'Lemon, it., P' 128
- Lindley (1965) 73
- loss function 11
- Lundqvist and Sjöberg (1969) 110
- Maritz (1966) 21, 26, 114
 - (1967) 21, 62
 - (1970) 14, 26, 27, 53, 71, 73, 103, 111, 114
- 'Martz' 129
- 'Martz, P' 129
- modified non-parametric EB estimator 46
- Monte Carlo simulations 22, 23, 25, 26, 31, 32, 35, 36, 99, 111
- naturally conjugate distributions 26
- non-parametric EB procedures v. EB procedures 39
- omitted element 39
- omitted indices 52
- optimal r 53, 66, 68, 113
- parameter values 13
- 'Parametric' 128
- 'Parametric, P' 128
- partition 51
- Parzen (1962) 26, 32
- past data 13
- prior distribution 11
- posterior distribution 12
- quadratic loss 11
- r -value 51, 111
- R_{n+1} -ratio 37, 100
- Raiffa and Schlaifer (1961) 26
- regression estimators 62
- risk 11
- Romberg's method 110
- Robbins (1955) 14, 17, 103
 - (1964) 15
- 'Robbins' 127
- 'Robbins, P' 127
- Rutherford and Krutchkoff (1969) 20
- S-factor 37, 100
- simultaneous probability function 40
- smooth 21, 62
- supplementary sample 30
- traditional estimator 15

